

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:04:33 ; Search time 137.122 Seconds
(without alignments)
3768.837 Million cell updates/sec

Title: US-09-836-602-6

Perfect score: 1509

Sequence: 1 MPRGTALCLLSLLPSGFM.....LVTSGALLAVLIGTGYFLMN 301

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+p2n.model -DEV=xlp
-Q/cqn2_1/USPTO_spool/US09836602/runat_29102002_101143_3758/app_query.fasta_1.1429
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602 @CGN_1.1.312 @runat_29102002_101143_3758 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_032802.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	1509	100.0	906	24	AAD22663	Human CD34 surface
2	1509	100.0	951	24	AAD22662	Human CD34 surface
3	1509	100.0	1122	24	AAD22661	Human full-length
4	1509	100.0	2463	22	AAH75132	Nucleotide sequenc
5	1509	100.0	2615	21	AAF20898	Human CD-34 polynu
6	1509	100.0	2615	21	AAA34776	Human adenosine re
7	1509	100.0	3490	21	AAF20899	Human CD-34 polynu
8	1509	100.0	3490	21	AAA34777	Human adenosine re
9	839	55.6	1065	18	AAT59508	Porcine CD34 from
10	508	33.7	615	21	AAC00249	Human secreted pro
11	311	20.6	233	22	ABA51134	Human breast cell
12	311	20.6	233	22	ABA69132	Human foetal liver
13	311	20.6	233	22	ABA36064	Probe #14530 for g
14	311	20.6	233	22	AAK17442	Human brain expres
15	311	20.6	233	22	AAK43242	Human bone marrow
16	311	20.6	233	22	AAI24018	Probe #13951 for g
17	311	20.6	233	22	AAI49317	Probe #18003 used
18	311	20.6	233	22	AAI09602	Probe #9593 used t
19	265	17.6	209	22	ABA50740	Human breast cell
20	265	17.6	209	22	ABA68710	Human foetal liver
21	265	17.6	209	22	ABA35674	Probe #14140 for g
22	265	17.6	209	22	AAK17050	Human brain expres
23	265	17.6	209	22	AAK42833	Human bone marrow
24	265	17.6	209	22	AAI23597	Probe #13530 for g
25	265	17.6	209	22	AAI48910	Probe #17596 used
26	265	17.6	209	22	AAI09212	Probe #9203 used t
27	233	15.4	235	22	AAS39236	Novel human diagno
28	170.5	11.3	406	22	ABA45618	Human breast cell
29	170.5	11.3	406	22	ABA56124	Human foetal liver
30	170.5	11.3	406	22	ABA25775	Probe #4241 for ge
31	170.5	11.3	406	22	AAK04311	Human brain expres
32	170.5	11.3	406	22	AAK29806	Human bone marrow
33	170.5	11.3	406	22	AAI14396	Probe #4329 for ge
34	170.5	11.3	406	22	AAI35770	Probe #4456 used t
35	170.5	11.3	406	22	AAI04219	Probe #4210 used t
36	159	10.5	13154	20	AAI13275	Enterococcus faeca
37	150.5	10.0	2336	23	ABL25662	Drosophila melanog
38	148	9.8	40875	18	AAT80043	Insert from cosmid
39	147.5	9.8	7720	21	AA53800	Genomic DNA encodi
40	145.5	9.6	4116	23	ABL24389	Drosophila melanog
41	145.5	9.6	5162	23	ABL20278	Drosophila melanog
42	145.5	9.6	6116	23	ABL24388	Drosophila melanog
43	144.5	9.5	8298	22	AAK72613	Human immune/haema
44	144	9.5	5163	19	AAV20700	Cryptosporidium pa
45	144	9.5	5163	21	AAA61849	ORF encoding a por

ALIGNMENTS

RESULT 1
AAD22663
ID AAD22663 standard; DNA; 906 BP.
XX
AC AAD22663;
XX
DT 26-FEB-2002 (first entry)
XX
DE Human CD34 surface antigen deleted variant (dCD34) encoding DNA.
XX
KW Human; surface marker; surface antigen; T lymphocyte; gene therapy;
XX CD34 deleted variant; dCD34; ds.
OS Homo sapiens.
OS Synthetic.
XX
FH Key
FT CDS Location/Qualifiers
1..906

FT /*tag= a
 FT /product= "Human CD34 surface antigen
 FT deleted variant (dCD34)"
 XX

PN EP1148066-A1.

XX 24-OCT-2001.

XX 18-APR-2001; 2001EP-0109374.

XX 18-APR-2000; 2000DE-1019075.

XX (ZAND/) ZANDER A R.

XX Zander AR;

XX WPI; 2002-019289/03.

DR P-PSDB; AAE13543.

XX New gene transfer vector (accession number DSM13396) containing a
 PT transgene and a nucleic acid sequence coding for a surface marker.
 PT useful in gene therapy, and for detecting genetically modified cells or
 PT cells which do not express CD34 -

XX Claim 3; Page 14-15; 28pp; English.

XX The patent discloses a gene transfer vector (accession number DSM13396)
 CC containing a transgene and a nucleic acid sequence coding for a surface
 CC marker. The surface marker is the CD34 surface antigen, its fragment or
 CC variant. The vector is useful for in vitro transduction of T lymphocytes,
 CC for gene therapy, and in the enrichment, detection and analysis of cells
 CC in vitro that do not naturally express CD34. T lymphocytes transduced
 CC with the vector are also useful in gene therapy. The CD34 nucleic acid
 CC sequences (marker genes), their fragments or variants are used for
 CC detecting genetically modified cells or cells which do not naturally
 CC express CD34. The present sequence is a DNA encoding human CD34 surface
 CC antigen deleted variant (dCD34).

XX Sequence 906 BP; 247 A; 250 C; 199 G; 210 T; 0 other;

XX Alignment Scores:

Pred. No.: 2.47e-110 Length: 906
 Score: 1509.00 Matches: 301
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 24 Gaps: 0

US-09-836-602-6 (1-301) x AAD22663 (1-906)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
 Db 1 ATGCGGGGGCTGGACCGCGCTTTGCTGCTGAGTTGTCGCTCTTGGGTTTCATGAGT 60
 Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 Db 61 CTTGACAAACACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTCACAAAGTT 120
 Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
 Db 121 TCTACAATGTATCTCTACCAAGAACTACACACCTAGTACCTTGGAAAGTACCAAGCTG 180
 Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValIys 80
 Db 181 CACCCTGTGCTCAACATGGCAATGAGCCACAAACATACAGAAACGACAGTCAAA 240
 Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 Db 241 TTCACATCTACTCTGTGATACCTCAGTTATGGAACACAAACTCTTCTGTCCAGTCA 300
 Qy 101 GlnThrSerValIleSerThrValPheThrThrThrProAlaAsnValSerThrProGluThr 120
 Db 301 CAGACCTCTGTGTAATCAGCACAGTGTTCACCAACCCCAACGTTTCAACTCCAGAGACA 360

Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrThrSer 140
 Db 361 ACCTTGAGAGCCTAGCCTGTCACCTGGAAATGTTTCAGACCTTTCAACACCTAGCAGTACG 420
 Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
 Db 421 CTTGCAACATCTCCCACTAAACCCCTATACATCATCTCTCTCTATCTTAAAGTGACATCAAG 480
 Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
 Db 481 GCAGAAATCAAAATGTTTCAGGCATCAGAGAGTGAATTTGACTCAGGGCATCTGCCTGGAG 540
 Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
 Db 541 CAAAATAGACCTCCAGCTGTCGGAGTTTAAAGAGACAGGGAGAGGGCTGGCCCGA 600
 Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
 Db 601 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGCTGGGGCCAGGTATGCTCCCTGCTC 660
 Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
 Db 661 CTTGCCAGTCTGAGTGGAGCCCTCAGTGTCTACTGTGCTTGGCCAAACAGAACAGAA 720
 Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
 Db 721 ATTTCCAGCAACTCCCACTTATGAAAAGCACCACCAATCTGACCTGAAAAAGCTGGGATC 780
 Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
 Db 781 CTAGATTTCACTGACGCAAGATGTTGCAAGCCACCAAGCTATTCACCAAGACCCCTGATT 840
 Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
 Db 841 GCATGTGTCACTCGGAGCCCTGCTGCTGTCTTGGGCATCACTGGCTATTTCTCTGATG 900
 Qy 301 Asn 301
 Db 901 AAT 903
 RESULT 2
 AAD22662
 ID AAD22662 standard; DNA; 951 BP.
 XX
 AC AAD22662;
 XX
 DT 26-FEB-2002 (first entry)
 XX
 DE Human CD34 surface antigen truncated variant (tCD34) encoding DNA.
 KW Human; surface marker; surface antigen; T lymphocyte; gene therapy;
 KW CD34 truncated variant; tCD34; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..951
 FT FT /*tag= a
 FT FT /product= "Human CD34 surface antigen
 FT FT truncated variant (tCD34)"
 XX
 PN EP1148066-A1.
 XX
 PD 24-OCT-2001.
 XX
 PF 18-APR-2001; 2001EP-0109374.
 XX
 PR 18-APR-2000; 2000DE-1019075.
 XX
 PA (ZAND/) ZANDER A R.
 XX

PI Zander AR;
 XX WPI; 2002-019289/03.
 DR P-PSDB; AAE13542.
 XX New gene transfer vector (accession number DSM13396) containing a
 PT transgene and a nucleic acid sequence coding for a surface marker,
 PT useful in gene therapy, and for detecting genetically modified cells or
 PT cells which do not express CD34 -
 XX
 PS Claim 3; Page 12-13; 28pp; English.
 XX The patent discloses a gene transfer vector (accession number DSM13396)
 CC containing a transgene and a nucleic acid sequence coding for a surface
 CC marker. The surface marker is the CD34 surface antigen, its fragment or
 CC variant. The vector is useful for in vitro transduction of T lymphocytes,
 CC in gene therapy, and in the enrichment, detection and analysis of cells
 CC in vitro that do not naturally express CD34. T lymphocytes transduced
 CC with the vector are also useful in gene therapy. The CD34 nucleic acid
 CC sequences (marker genes), their fragments or variants are used for
 CC detecting genetically modified cells or cells which do not naturally
 CC express CD34. The present sequence is a DNA encoding human CD34 surface
 CC antigen truncated variant (tCD34).
 XX
 SQ Sequence 951 BP; 261 A; 264 C; 213 G; 213 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,63e-110 Length: 951
 Score: 1509.00 Matches: 301
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 24 Gaps: 0
 US-09-836-602-6 (1-301) x AAD22662 (1-951)
 QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
 DB 1 ATGCCGCGGGCTGGACCGCGCTTTGCTGCTGAGTTTGCTGCCCTCTGGGTTTCATGAGT 60
 QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 DB 61 CTTGACAAACAGCGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAATGTT 120
 QY 41 SerThrAsnValSerTyrglnGluThrThrProSerThrLeuGlySerThrSerLeu 60
 DB 121 TCTCAATGTTATCTACCAAGAACTACACACCTAGTACCTCTGGAGTAGCACGCCG 180
 QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
 DB 181 CACCCTGTCTCAACATGGCAATGAGGCCACAAACATCACAGAAACGACAGTCAAA 240
 QY 81 PheThrSerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValGlnSer 100
 DB 241 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTCTCTCCAGTCA 300
 QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 DB 301 CAGACCTCTGTAAATCAGACAGTGTTCACCCAGCCAGCAAGTTTCACTCCAGAGACA 360
 QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
 DB 361 ACCTTTGAAGCCTAGCCTGTGACCTGGAAATGTTTTCAGACCTTTCAACCACTAGCACTAGC 420
 QY 141 LeuAlaThrSerProThrLysProTyrrThrSerSerSerProIleLeuSerAspIleLys 160
 DB 421 CTTGCAACATCTCCCACTAAACCCCTATACATCATCTCTCTCTATCTTCTAGTGCATCAAG 480
 QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
 DB 481 GCAGAAATCAATGTTAGGCATCAGAGAGTGAATGACTCAGGGCATCTGCCCTGGAG 540
 QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluLeuAlaArg 200
 DB 541 CAAATAAGACCTCCAGCTGTGCGAGTTTAAGAAGGACAGGGAGAGGGCTGGCCCGA 600
 QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
 DB 601 GTGCTGTGTGGGAGGAGCAGCTGATGCTGATGTGGGGCCCGAGGTATGCTCCCTGCTC 660
 QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu 240
 DB 661 CTTGCCAGTCTGAGGTGAGGCTCAGTGTCTACTGCTGCTGGCTTTGGCCCAACAGACAGAA 720
 QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
 DB 721 ATTTCCAGCAAACTCCAATTATGAAAAGCACCAATCTGACCTGAAAAGCTGGGGATC 780
 QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrrSerGlnLysThrLeuIle 280
 DB 781 CTAGATTTTCACTGACCAAGATGTTGCAAGCCACACAGAGCTATTTCCAAAAGACCTGATT 840
 QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrrPheLeuMet 300
 DB 841 GCACCTGGTCACCTCGGAGCCCTGCTGCTGCTGCTGGGATCCTGCTATTTCTCTGATG 900
 QY 301 Asn 301
 DB 901 AAT 903
 RESULT 3
 AAD22661
 ID AAD22661 standard; DNA; 1122 BP.
 AC AAD22661;
 XX
 DT 26-FEB-2002 (first entry)
 XX
 DE Human full-length CD34 (fICD34) surface antigen encoding DNA.
 KW Human; surface marker; surface antigen; T lymphocyte; gene therapy;
 full-length CD34; fICD34; ds.
 OS Homo sapiens.
 XX
 FT Key Location/Qualifiers
 CDS 1..1122 /*tag= a
 /product= "Human full-length CD34 surface antigen"
 XX
 PN EP1148066-A1.
 XX
 PD 24-OCT-2001.
 XX
 PF 18-APR-2001; 2001EP-0109374.
 XX
 PR 18-APR-2000; 2000DE-1019075.
 XX
 PA (ZAND/) ZANDER A R.
 XX
 PI Zander AR;
 XX
 DR WPI; 2002-019289/03.
 DR P-PSDB; AAE13541.
 XX
 PT New gene transfer vector (accession number DSM13396) containing a
 PT transgene and a nucleic acid sequence coding for a surface marker,
 PT useful in gene therapy, and for detecting genetically modified cells or
 PT cells which do not express CD34 -
 XX
 PS Claim 3; Page 9-11; 28pp; English.
 XX The patent discloses a gene transfer vector (accession number DSM13396)
 CC containing a transgene and a nucleic acid sequence coding for a surface
 CC marker. The surface marker is the CD34 surface antigen, its fragment or
 CC variant. The vector is useful for in vitro transduction of T lymphocytes,

CC for gene therapy, and in the enrichment, detection and analysis of cells
CC in vitro that do not naturally express CD34. T lymphocytes transduced
CC with the vector are also useful in gene therapy. The CD34 nucleic acid
CC sequences (marker genes), their fragments or variants are used for
CC detecting genetically modified cells or cells which do not naturally
CC express CD34. The present sequence is a DNA encoding human full-length
CC CD34 (f1CD34) surface antigen.
XX
SQ Sequence 1122 BP; 305 A; 311 C; 270 G; 236 T; 0 other;

Alignment Scores:
Pred. No.: 3,25e-110 Length: 1122
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 24 Gaps: 0

US-09-836-602-6 (1-301) x AAD22661 (1-1122)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
DB 1 ATGCCGGGGGCTGGAGCCGGCTTGTCTGCTGAGTTGCTCTCTGGGTTTCATAGT 60
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
DB 61 CTTGCACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 120
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
DB 121 TCTACAATGTATCTACCAAGAAGACTACAAACACTAGTACCCTTGGGAAGTACCAGCTG 180
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
DB 181 CACCCTGTGTCTCAACTGAGCAATGAGGCCACAAACATCACAGAAACGACAGTCAAA 240
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
DB 241 TTCACATCTACCTCTGTGATTAACCTCAGTTTATGGAACACAACTCTTCTGTCAGTCA 300
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
DB 301 CAGACCTCTCTAATCAGCAGAGTGTTCACCACCCAGCCCAACGTTTCAACTCCAGAGACA 360
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
DB 361 ACCTTGAAGCCTAGCCGTGTCACCTGGAAATGTTTCAGACCTTTCACCACTAGCCTAGC 420
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLys 160
DB 421 CTTGCAACATCTCCCACTAAACCTTATACATCATCTCTCTATCTAAGTGACATCAAG 480
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
DB 481 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATGACTCAGGGCATCTGCTGGAG 540
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
DB 541 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAGACAGGGGAGGGCCCTGGGCCGA 600
QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlnValCysSerLeuLeu 220
DB 601 GTGCTGTGTGGGAGGAGGAGGCTGATGCTGATGCTGGGGCCAGGTATGCTCCCTGCTC 660
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
DB 661 CTTGCCCCAGTCTCAGGTGAGGCCCTCAGTGTCTACTGCTGCTTGTGGCCAAACAGAGAA 720
QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
DB 721 ATTTTCAGCAAACTCCAACTTATGAAGAAGACCACTACTGACCTGAAAGAGCTGGGGATC 780
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280

DB 781 CTAGATTTCACCTGAGCAAGATGTTGCAAGCCACACAGAGCTATTCCCAAAAGACCCCTGATT 840
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
DB 841 GCACCTGTGCTACCTCGGAGCCCTGCTGGCTGTCTTGGGCATCACTGGCTATTTCCTGATG 900
QY 301 Asn 301
DB 901 AAT 903
RESULT 4
AAH75132
ID AAH75132 standard; DNA: 2463 BP.
XX
AC AAH75132;
XX
DT 13-NOV-2001 (first entry)
XX
DE Nucleotide sequence of a human CD34 polypeptide.
XX
KW Human; CD34 gene; blast crisis; chronic myelogenous leukemia;
KW nm23-H4 kinase gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 91..1248
FT /tag= a
FT /product= "CD34"
XX
PN WO200164946-A1.
XX
PD 07-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-JP01485.
XX
PR 02-MAR-2000; 2000JP-0058043.
XX
PA (TAKI) TAKARA SHUZO CO LTD.
XX
PI Mano H, Miyazato A, Ueno S, Yoshida K, Yamanaka T, Ikeda U;
PI Shimada K, Hatake K, Ozawa K, Asada K, Kato I;
XX
DR WPI: 2001-550191/61.
DR P-PSDB; AAG67120.
XX
PT Method for detecting chronic myelogenous leukemia by comparing
PT expression levels of CD34 and nm23-H4 genes -
XX
PS Disclosure; Page 36-40; 60pp; Japanese.
XX
CC The present sequence encodes a human CD34 polypeptide. The
CC specification describes a method of detecting blast crisis in chronic
CC myelogenous leukemia. The method comprises comparing the amounts of
CC expression of at least two genes in a sample, particularly CD34 gene
CC and nm23-H4 kinase gene. The method allows the worsening stages of
CC chronic myelogenous leukemia to be easily detected at a high
CC reliability.
XX
SQ Sequence 2463 BP; 589 A; 709 C; 580 G; 585 T; 0 other;
Alignment Scores:
Pred. No.: 9e-110 Length: 2463
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0
US-09-836-602-6 (1-301) x AAH75132 (1-2463)
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20

KW	Low adenosine oligonucleotide; phosphorothioate; allergy;
KW	human; airway disorder; bronchoconstriction; lung inflammation;
KW	surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW	immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW	respiratory obstruction; pulmonary obstruction; impeded expiration;
KW	surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW	respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis
KW	pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW	chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW	cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO2000062736-A2.
XX	
PD	26-OCT-2000.
XX	
PF	24-MAR-2000; 2000WO-US08020.
XX	
PR	06-APR-1999; 99US-0127958.
XX	
PA	(UVEC-) UNIV EAST CAROLINA.
PA	(NYCE/) NYCE J W.
XX	
PI	Nyce JW;
XX	
DR	WPI; 2000-679539/66.
XX	
PT	Low adenosine (A) content antisense oligonucleotides which do not
PT	trigger adenosine receptors during metabolism, useful e.g. for treating
PT	cancers and respiratory obstructions -
XX	
PS	Disclosure; Page 292-293; 1592pp; English.
XX	
CC	The present invention describes low adenosine (A) content antisense
CC	oligonucleotides and compositions (I) comprising them. In the antisense
CC	oligonucleotides the A is replaced by a 'Universal' or alternative base
CC	(I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC	immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC	The antisense oligonucleotides and (I) can be used to down-regulate the
CC	expression and or activity of target polypeptides associated with
CC	lung/respiratory disorders and malignancies, such as stimulating and
CC	activating peptide factors and transmitters, transcription factors,
CC	immunoglobulins and antibodies, antibody receptors, cytokines and
CC	chemokines, endogenously produced specific and non-specific enzymes,
CC	binding proteins, adhesion molecules and their receptors, cytokine and
CC	chemokine receptors, adenosine receptors, bradykinin receptors, central
CC	nervous system (CNS) and peripheral nervous and non-nervous system
CC	receptors, CNS and peripheral nervous and non-nervous system peptide
CC	transmitters, defensins, growth factors, vasoactive peptides and
CC	receptors, binding proteins and malignancy associated proteins. The
CC	antisense oligonucleotides may be used in this way to treat disorders
CC	including respiratory obstruction (especially pulmonary obstruction
CC	and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC	and/or surfactant hypoproduction which are associated with a disease or
CC	condition selected from pulmonary vasoconstriction, inflammation,
CC	allergies, asthma, impeded respiration, respiratory distress syndrome
CC	(RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC	hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC	pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC	and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC	fragments and antisense oligonucleotides used in the exemplification of
CC	the present invention.

Alignment Scores:	9.72e-110	Length:	2615
Pred. No.:	1509.00	Matches:	301
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	100.00%	Gaps:	0
DB:	21		

XX	AAF20898;
XX	
XX	14-MAR-2001 (first entry)
DT	
XX	Human CD-34 polynucleotide fragm
DE	
XX	

us-09-836-602-6 (1-301) x AAF20898 (1-2615)

```
Qy 1 MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 294 ATGCCGGGGCTGAGACCGCTTTCGTGCTGAGTTGCTGCGCTTCGGGTTCATGAGT 353
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 354 CTTCACACACACGGTACTGCTACCCAGAGTACCTACCCAGGGAGACATTTTCAATGTT 413
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 414 TCTACAAATGTATCTACCAAGAACTACAACACCTAGTACCTTGGAAGTACAGCCTG 473
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVallys 80
Db 474 CACCTGTGTCTCAACATGGCAATGAGGCCACAACAACATCACAGAAACACAGTCAAA 533
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 534 TTCACATCTACTCTGTGATAACTCAGTTTATGGAAACACAAACTCTTCGTCCAGTCA 593
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 594 CAGACCTCTGTAATCAGCACAGTGTTCACCACCCAGCCCAACGTTTCAACTCCAGAGACA 653
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 654 ACCTTGAAGCCTAGCCTGTACCTCGGAATGTTTCAGACCTTTCAACACACAGTACAGT 713
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLys 160
Db 714 CTTGCAACATCTCCCACTAAACCCATACATCATCTTCTCTATCTTAAGTGACATCAAG 773
Qy 161 AlaGluLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 774 GCAGAAATCAAAATGTTCAGGCATCAGAGAGTGAATTTGACTCAGGGCATCTGCTGGAG 833
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 834 CAAATATAGACCTCCAGCTGTGCGGAGTTTAAGAGGACAGGGAGAGGGCTGGCCGA 893
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 894 GTGCTGTGGGGAGGAGCAGCTGATGCTGCTGGGGCCAGGTATGCTCCCTGCTC 953
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 954 CTTGCCAGTCTGAGGTGAGGCCTCAGTGTCTACTGTGCTTTGGCCAAACAGAACAGAA 1013
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLeuGlyIle 260
Db 1014 ATTTCCAGCAACCTCCAACTTATGAAGAGCACCACCACTGACCTGAAAGCTGGGATC 1073
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 1074 CTAGATTTCAGTACGCAAGATGTTTGAAGCCACAGAGCTATTCCCAAAAGACCTGATT 1133
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 1134 GCATGTGTACCTCGGGAGCCCTGCTGGCTGTCTTGGGCATCACTGGCTATTTCCTGATG 1193
Qy 301 Asn 301
Db 1194 AAT 1196
```

RESULT 6
AAA34776
ID AAA34776 standard; DNA: 2615 BP.
XX
AC
AC
XX

28-JUL-2000 (first entry)
Human adenosine receptor related polynucleotide SEQ ID NO:2465.
Human; adenosine receptor; low adenosine antisense oligonucleotide; phosphorothioate; impaired respiration; inflammation; allergy; allergic disease; bronchoconstriction; inhibitor; antiinflammatory; antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway; lung disease; ischaemic condition; pulmonary vasoconstriction; asthma; respiratory distress syndrome; pain; cystic fibrosis; emphysema; pulmonary hypertension; chronic obstructive pulmonary disease; COPD; cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
Homo sapiens.
WO200009525-A2.
24-FEB-2000.
03-AUG-1999; 99WO-US17712.
03-AUG-1998; 98US-0095212.
(UYEC-) UNIV EAST CAROLINA.
Nyce JW;
WPI; 2000-205971/18.
New antisense oligonucleotides useful for treating e.g. pulmonary vasoconstriction, inflammation, allergies, asthma, hypertension, bronchitis, emphysema, respiratory distress syndrome, ischemia or cancers -
Disclosure; Page 602; 1343pp; English.
The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, e.g. impaired airflow, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma, impaired respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas, carcinomas, and cancers which may metastasize to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.
SQ Sequence 2615 BP; 616 A; 763 C; 627 G; 609 T; 0 other;

Alignment Scores:
Pred. No.: 9.72e-110 Length: 2615
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-836-602-6 (1-301) x AAA34776 (1-2615)

```
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 294 ATCCCGGGGGTGGACCGGCTTGGCTGCTGAGTTTGGCTTCTGGGTTTCATGAGT 353
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 354 CTTGACAAACAGGTACTGCTACCCAGAGATTACCTACCCAGGGAACATTTTCAAAATGTT 413
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 414 TCACAAATGTATCCTACCAAGAACTACCAACACCTAGTAGTCCCTTGGAGTACCAGCGTG 473
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 474 CACCCTGTCTCAACATGCCAATGAGGCCACCAACAACATCAGAAAGACAGACGTCAA 533
QY 81 PheThrSerThrSerValIleThrSerValThrGlyAsnThrAsnSerSerValGlnSer 100
Db 534 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 593
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 594 CAGACCTCTGTATACGACACAGTGTTCACCACCCCAAGCCAGCTTCAACTCCAGAGACA 653
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 654 ACCTTGAAGCCTAGCCTGTCCACCTGGAATGTTTCAGACCTTTCACACCATGACCTAGC 713
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 714 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTCCTAAAGTGCATCAAG 773
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 774 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAAATTTGACTCAGGGGCATCTGCCTGGAG 833
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlyLeuAlaArg 200
Db 834 CAANAATGACCTCCACCTGTGGAGTTTAAGNAGGACAGGGAGGCGCTGGGCCCGA 893
QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlnValCysSerLeuLeu 220
Db 894 GTGCTGTGTGGGAGGACAGGCTGATGCTGATGCTGGGGCCCAAGGTATGCTCCCTGCTC 953
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 954 CTTGCCCAGTCTGAGGTGAGGCGCTCACTGTCTACTGTGTCTTGGCCAAACAGAACAGAA 1013
QY 241 IleSerSerLysLeuGlnLeuMetLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 1014 ATTTCCAGCAAACTCCAACTTATGAAAACAGCACCAATCTGACCTGAAAAGCTGGGGATC 1073
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 1074 CTAGATTCTACTGACGAAGATGTTGCAAGCCACCAGAGCTATTCCCAAAAGACCCCTGATT 1133
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 1134 GCACGTGCTACCTGGGAGGCCCTGCTGGCTGCTTGGGCATCACTGGCTATTTCCTGATG 1193
QY 301 Asn 301
Db 1194 AAT 1196
RESULT 7
AAF20899
ID AAF20899 standard; DNB: 3490 BP.
XX AAF20899;
AC AAF20899;
XX
DT 14-MAR-2001 (first entry)
XX
```

Human CD-34 polynucleotide fragment #2466.

Low adenosine antisense oligonucleotide; phosphorothioate; allergy; human; airway disorder; bronchoconstriction; lung inflammation; surfactant depletion; respiratory; bronchodilator; antiinflammatory; immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic; respiratory obstruction; pulmonary obstruction; impeded respiration; surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS; respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis; pulmonary hypertension; emphysema; pulmonary transplantation rejection; chronic obstructive pulmonary disease; pulmonary infection; bronchitis; cancer; ss.

Homo sapiens.

WO200062736-A2.

26-OCT-2000.

24-MAR-2000; 2000WO-US08020.

06-APR-1999; 99US-0127958.

(UYEC-) UNIV EAST CAROLINA.

(NYCE/) NYCE J W.

Nyce JW;

WPI; 2000-679539/66.

Low adenosine (A) content antisense oligonucleotides which do not trigger adenosine receptors during metabolism, useful e.g. for treating cancers and respiratory obstructions -

Disclosure; Page 291-292; 1592pp; English.

The present invention describes low adenosine (A) content antisense oligonucleotides and compositions (I) comprising them. In the antisense oligonucleotides the A is replaced by a 'universal' or alternative base. (I) can have respiratory, bronchodilator, antiinflammatory, analgesic, immunosuppressive, antiasthmatic, hypotensive and cytostatic activities. The antisense oligonucleotides and (I) can be used to down-regulate the expression and or activity of target polypeptides associated with lung/respiratory disorders and malignancies, such as stimulating and activating peptide factors and transmitters, transcription factors, immunoglobulins and antibodies, antibody receptors, cytokines and chemokines, endogenously produced specific and non-specific enzymes, binding proteins, adhesion molecules and their receptors, cytokine and chemokine receptors, adenosine receptors, bradykinin receptors, central nervous system (CNS) and peripheral nervous and non-nervous system receptors, CNS and peripheral nervous and non-nervous system peptide transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation, allergies, asthma, impeded respiration, respiratory distress syndrome (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), pulmonary transplantation rejection, pulmonary infections, bronchitis, and/or cancer. AAF18434 to AAF21543 represent human polynucleotide fragments and antisense oligonucleotides used in the exemplification of the present invention.

Sequence 3490 BP; 880 A; 966 C; 872 G; 772 T; 0 other;

Alignment Scores:

Pred. No.:	1.41e-109	Length:	3490
Score:	1509.00	Matches:	301
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

AC	AAA34777;
XX	28-JUL-2000 (first entry)
XX	Human adenosine receptor related polynucleotide SEQ ID NO:2466.
DE	Human; adenosine receptor; low adenosine antisense oligonucleotide; phosphorothioate; impaired respiration; inflammation; allergy; allergic disease; bronchoconstriction; inhibitor; antiinflammatory; antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway; lung disease; ischaemic condition; pulmonary vasoconstriction; asthma; respiratory distress syndrome; pain; cystic fibrosis; emphysema; pulmonary hypertension; chronic obstructive pulmonary disease; COPD; cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX	Homo sapiens.
OS	
XX	WO200009525-A2.
PN	24-FEB-2000.
XX	
PD	
XX	03-AUG-1999; 99WO-US17712.
PF	
XX	03-AUG-1998; 98US-0095212.
PR	(UYEC-) UNIV EAST CAROLINA.
XX	
PA	Nyce JW;
PI	
XX	WPI; 2000-205971/18.
DR	
XX	New antisense oligonucleotides useful for treating e.g. pulmonary vasoconstriction, inflammation, allergies, asthma, hypertension, or bronchitis, emphysema, respiratory distress syndrome, ischemia or cancers -
PT	
XX	Disclosure; Page 603; 1343pp; English.
PS	
XX	The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas, carcinomas, and cancers which may metastasize to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185 sequences are also called SEQ ID NO:1 to 185, and then the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.
CC	
XX	Sequence 3490 BP; 880 A; 966 C; 872 G; 772 T; 0 other;
SQ	
Alignment Scores:	
Pred. No.:	1 41e-109 Length: 3490
Score:	1509.00 Matches: 301
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
Gap:	21 Gaps: 0
DR:	

ID AAT59508 SCIN0010; CDNA; 1063 BP.
XX
XX
AC AAT59508;
XX

QY 33 -----ThrGInGlyThrPheSerAsn----- 39
DB 184 GCGCGCGCTACCGCAGCGCCAGCTATTCACGACATCTTTCACACT

```
QY 40 ValSerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSer 59
Dbb 244 GTTCTTACAAATATATCCAAACAGGAAACACATCA--GATGCTTTCGAAAGTCCAGC 300
QY 60 LeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVal 79
Dbb 301 CTCACACTGTCTCTCAGGCGCAGCAGTGGGACCACTAGCCATCTCAGCCCTACAGTT 360
QY 80 LysPheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGln 99
Dbb 361 AATTTCATGCTACCTCGCGGTCACCTCGTCCCGAAACCGTTAACTCTCTGTCGACG 420
QY 100 SerGlnThrSerValIleThrSerValPheThrThrProAlaAsnValSerThrProGlu 119
Dbb 421 CCTCAGACCTCT--CTAGCCACAGCGTCTCCGCCACCATCACTTTACAACTTCAGAG 477
QY 120 ThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThr 139
Dbb 478 GTGACCTCGACGCCACGACGTTCCCGAGGAATGTTTCAGACCCCTCTACAAACAGTACC 537
QY 140 SerLeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIle 159
Dbb 538 AGCCTGCGAGATCCCCACAGCCCTACACATCATCTCTCCCTACCCAGGTAGCCAC 597
QY 160 LysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeu 179
Dbb 598 AAGGGGGAAGTCAAAATGTGCCAAATCAAGAGGTGAAATTCACCAAGGTATCTGCCGTG 657
QY 180 GluGlnAspLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla 199
Dbb 658 GAGCAAAATGAGACCTCCCGCTGCGGAGAGTTTAAGAGGCAATGAGAGAGTGTATG 717
QY 200 ArgValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeu 219
Dbb 718 CAATCTGTGTGGCAGCAGCAGGTGAGCGCGGCCAGGG-----GTGTGCTCTGTG 771
QY 220 LeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThr 239
Dbb 772 CTCCTTGCCCAATCTGAGTGAACCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 831
QY 240 GluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLeuGly 259
Dbb 832 GAACCTAGCAGCAAGTTCCTGCTCTGGAAGAACCCAGCTGGAACCTGAGAGATGAGC 891
QY 260 IleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeu 279
Dbb 892 ATCCAAACTTCTCGAAACAGATGTTAGGACCCAGCAGACTACTCCCGAAAGACCTTG 951
QY 280 IleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeu 299
Dbb 952 ATTGCACTGGTCACTCGGGGATCCTGCTGGCTGTCTTGGGCATCACTGGCTACTTGCTG 1011
QY 300 MetAsn 301
Dbb 1012 ATGAAC 1017
RESULT 10
AAC00249
ID AAC00249 standard; cDNA; 615 BP.
XX
AC AAC00249;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 247.
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
OS Homo sapiens.
PN EP1033401-A2.
```

```
XX 06-SEP-2000.
PD
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI; 2000-500381/45.
XX P-PSDB; AAG00243.
DR
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PS Claim 1; SEQ ID 247; 7lpp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 615 BP; 149 A; 181 C; 159 G; 119 T; 7 other;
```

```
Alignment Scores:
Pred. No.: 1,7e-31 Length: 615
Score: 508.00 Matches: 99
Percent Similarity: 96.12% Conservative: 0
Best Local Similarity: 96.12% Mismatches: 4
Query Match: 33.66% Indels: 0
DB: 21 Gaps: 0
```

US-09-836-602-6 (1-301) x AAC00249 (1-615)

```
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Dbb 305 ATGCGCGGGGCTGGACCGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 364
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Dbb 365 CTTGACAAACAGGTACTGTACCCCGAGAGTTACCTACCCGAGGAACATTTTCAAATGTT 424
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Dbb 425 TCTACAAATGATCTCTACCAAGAAACTACACACCTGTGACCTTGGAGAGTACCAGCTTG 484
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Dbb 485 CACCCTGTGCTCAACATGGCAATGAGGCCACAAACATCACAGAAACGACAGTSSNG 544
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Dbb 545 TTCACATCTASCTGTGTGATAASCTCAGTTTATGGRACACAAACTCTTCTGTCCAGTCA 604
QY 101 GlnThrSer 103
Dbb 605 CAGACCTCT 613
RESULT 11
ABA51134/c
```

ID ABA51134 standard; DNA; 233 BP.
 XX
 AC ABA51134;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human breast cell single exon nucleic acid probe #9829.
 XX
 KW Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00662.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-496933/54.
 XX
 PT New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 XX
 PS Claim 4; SEQ ID NO 9829; 327pp + sequence listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and Br 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
 Pred. No.: 1.76e-16 Length: 233
 Score: 311.00 Matches: 60
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 20.61% Indels: 0
 DB: 22 Gaps: 0

US-09-836-602-6 (1-301) x ABA51134 (1-233)

Qy 16 SerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGly 35
 Db TCTGGGTTTCATGAGTCTTGACAAACACGGTACTGTACCCAGAGTTACCTACCCAGGGA 172

Qy 36 ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeu 55
 Db ACATTTTCAAAATGTTTACAAATGTATCCTACCAGAACTACAACTAGTACCTT 112
 Qy 56 GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr 75
 Db 111 GGAAGTACCAGCTGCACCTGTGTCTCAACATGCAATGAGGCCACAAACAATCACA 52
 RESULT 12
 ABA69132/c
 ID ABA69132 standard; DNA; 233 BP.
 XX
 AC ABA69132;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human foetal liver single exon nucleic acid probe #17437.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX
 PS Claim 4; SEQ ID NO 17437; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
 Pred. No.: 1.76e-16 Length: 233
 Score: 311.00 Matches: 60
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 20.61% Indels: 0
 DB: 22 Gaps: 0

US-09-836-602-6 (1-301) x ABA69132 (1-233)

Qy 16 SerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGly 35
 Db TCTGGGTTTCATGAGTCTTGACAAACACGGTACTGTACCCAGAGTTACCTACCCAGGGA 172

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 30, 2002, 08:08:13 ; Search time 38.6131 Seconds
(without alignments)
1914.781 Million cell updates/sec

Title: US-09-836-602-6
Perfect score: 1509
Sequence: 1 MPRGWTALCLSLPSGFMS.....LVTSGALLVLGITGVFLMN 301

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlip
-Q=/cgn2_1/USPTO.spool/US09836602/runat_29102002_101145_3837/app_query.fasta_1.1429
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdd
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602_@cgnl_1_44_@runat_29102002_101145_3837 -NCPU=6 -ICPU=3
-NO_TLXPAY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	839	55.6	1065	2	US-08-475-634D-18
2	144	9.5	5163	3	US-08-700-651-1
3	144	9.5	5163	3	US-08-928-361B-4
4	144	9.5	5318	3	US-08-700-651-2
5	144	9.5	5318	3	US-08-928-361B-3
6	137.5	9.1	5511	3	US-08-928-361B-2
7	137.5	9.1	7334	3	US-08-928-361B-1
c 8	129	8.5	1505	1	US-07-915-246-1
9	128	8.5	390	4	US-09-197-649-7
c 10	127.5	8.4	3337	1	US-08-072-610-1
c 11	127.5	8.4	3337	2	US-08-719-822B-1
c 12	127.5	8.4	3337	4	US-09-092-458-1

13	125.5	8.3	2589	1	US-08-325-267A-3	Sequence 3, Appli
14	125.5	8.3	2685	3	US-08-362-525-21	Sequence 21, Appl
15	125.5	8.3	4614	1	US-08-325-267A-1	Sequence 1, Appl
16	123	8.2	2754	1	US-08-270-076A-10	Sequence 10, Appl
17	121	8.0	2214	6	5258502-1	Patent No. 5258502
18	121	8.0	3168	4	US-09-165-239A-3	Sequence 3, Appli
19	119.5	7.9	688	4	US-08-998-416-915	Sequence 915, App
20	119.5	7.9	1107	2	US-08-991-300-1	Sequence 1, Appli
c 21	116	7.7	2793	1	US-08-209-747-1	Sequence 1, Appli
c 22	116	7.7	2793	1	US-08-458-298-1	Sequence 1, Appli
23	113.5	7.7	2032	4	US-09-241-581B-5	Sequence 5, Appli
24	115.5	7.7	2032	5	PCT-US95-07721-5	Sequence 5, Appli
25	115	7.6	2584	3	US-08-758-662-8	Sequence 8, Appli
26	114.5	7.6	2093	1	US-08-287-001A-1	Sequence 1, Appli
27	114.5	7.6	2093	5	PCT-US95-09941-1	Sequence 1, Appli
28	113	7.5	36519	3	US-08-923-137-2	Sequence 2, Appli
29	110.5	7.3	2164	4	US-08-760-615-3	Sequence 3, Appli
30	108.5	7.2	4108	4	US-08-981-729-8	Sequence 8, Appli
31	108.5	7.2	4108	4	US-08-981-446B-1	Sequence 1, Appli
32	108	7.2	2150	2	US-08-861-464-13	Sequence 13, Appl
33	108	7.2	2150	2	US-08-396-001-13	Sequence 13, Appl
34	108	7.2	2150	4	US-09-323-433A-13	Sequence 13, Appl
c 35	108	7.2	9636	1	US-08-323-170B-1	Sequence 1, Appli
c 36	108	7.2	9636	4	US-08-954-441-1	Sequence 1, Appli
37	107.5	7.1	3833	1	US-08-917-320-18	Sequence 18, Appl
38	107.5	7.1	3833	5	PCT-US95-04611A-18	Sequence 18, Appl
39	107.5	7.1	5931	3	US-08-783-774-1	Sequence 1, Appli
40	106	7.0	10803	3	US-09-080-044-1	Sequence 1, Appli
41	106	7.0	19227	3	US-09-090-793-13	Sequence 13, Appl
42	106	7.0	40138	3	US-09-090-793-12	Sequence 12, Appl
43	105	7.0	3141	2	US-08-658-665-66	Sequence 66, Appl
44	105	7.0	3141	4	US-08-796-101-30	Sequence 30, Appl
45	105	7.0	3141	4	US-09-085-273-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1
US-08-475-634D-18
; Sequence 18, Application US/08475634D
; Patent No. 5962644
; GENERAL INFORMATION:
; APPLICANT: HAWLEY, Robert J.
; APPLICANT: MONROY, Rodney L.
; TITLE OF INVENTION: Antibodies to Porcine CD34 Positive
; TITLE OF INVENTION: and Uses Therefor
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS 6.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475.634D
; FILING DATE: June 7, 1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 61750-141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 973-994-1700
; TELEFAX: 973-994-1744
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:

002 UHAC11AGCAGUCAGG11CC1U6C11C1U6UHHAAAGCACCAAGTCTUGAAC1GAGAGAGAGATGAGC 899

QY 141 uAlaThrSerPr

002 UHAC11AGCAGUCAGG11CC1U6C11C1U6UHHAAAGCACCAAGTC1GAAAC1GAGAGAGAGATGAGC 089

QY 141 uAlaThrSerProThrLysProThrLysProThrSerSerSer 153
Db 973 CACCACAAACAACAACAACAACAACAACAACAACAACA 1009
RESULT 5
US-08-928-361B-3
; Sequence 3, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5318 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-928-361B-3
Alignment Scores:
Pred. No.: 6.04e-05 Length: 5318
Score: 144.00 Matches: 37
Percent Similarity: 47.37% Conservative: 26
Best Local Similarity: 27.82% Mismatches: 69
Query Match: 9.54% Indels: 1
DB: 3 Gaps: 0
US-09-836-602-6 (1-301) x US-08-928-361B-3 (1-5318)
QY 22 AspAsnAsnGly-ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSe 41
Db 612 GACA 671
QY 41 rThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeuHi 61
Db 672 GACA 731
QY 61 sProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLysPh 81
Db 732 AAGCACTAC 791
QY 81 eThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSerGl 101

Db 792 TACA 851
QY 101 nThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThrTh 121
Db 852 AACCAACA 911
QY 121 rLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSerLe 141
Db 912 CAAGAACA 971
QY 141 uAlaThrSerProThrLysProThrLysProThrSerSerSer 153
Db 972 CACCACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1008
RESULT 6
US-08-928-361B-2
; Sequence 2, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-928-361B-2
Alignment Scores:
Pred. No.: 0.000326 Length: 5511
Score: 137.50 Matches: 61
Percent Similarity: 39.19% Conservative: 46
Best Local Similarity: 22.34% Mismatches: 135
Query Match: 9.11% Indels: 31
DB: 3 Gaps: 7
US-09-836-602-6 (1-301) x US-08-928-361B-2 (1-5511)
QY 26 ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnValSer 45

Qy 204 GlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeuLeuAlaGln 223
Db 3334 ATCGAAAGATGACGATGATGAAATGGAATGCAATTTACAAATGATTCACAAATGATGAC 3393
Qy 224 SerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGluIle----- 241
Db 3394 ACGCATGTCGCTTCAGATTTAAAGGTTAAAGATGATGGGACACTATTTTCAGTAAGATGC 3453
Qy 242 -----SerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeu 258
Db 3454 GAAAGAGTGCAGTAACACTCCGATTCACAGATAGAGT----- 3492
Qy 259 GlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThr 278
Db 3493 -----TTGGATTTCACAATTCCTCCAGTAGCTGCGCATACAGCTGTTC----- 3537
Qy 279 LeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaVal 291
Db 3538 ATAATAGTTGGTGTGAGCGCGGTGGAAAAATTCACGTA 3576
RESULT 8
US-09-836-602-6 (1-301) x US-07-915-246-1 (1-1505)
; Sequence 1, Application US/07915246
; Patent No. 5401836
; GENERAL INFORMATION:
; APPLICANT: Baszczynski, Chris L.
; APPLICANT: Fallis, Lynne
; APPLICANT: Bellmare, Guy
; APPLICANT: Bolvin, Rodolphe
; TITLE OF INVENTION: A BRASSICA REGULATORY SEQUENCE FOR
; TITLE OF INVENTION: ROOT-SPECIFIC OR ROOT-ABUNDANT GENE EXPRESSION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McAndrews, Held, and Malloy
; STREET: 500 W. Madison St. 34th Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/915,246
; FILING DATE: 19920716
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 91 P 1125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 707-8889
; TELEFAX: 312 707-9155
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1505 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Brassica napus
; STRAIN: Westar
; DEVELOPMENTAL STAGE: Somatic
; TISSUE TYPE: Root
US-07-915-246-1

Alignment Scores: 0.000289 Length: 1505
Pred. No.:

Score: 129.00 Matches: 47
Percent Similarity: 48.78% Conservative: 33
Best Local Similarity: 28.66% Mismatches: 70
Query Match: 8.55% Indels: 14
DB: 1 Gaps: 5
US-09-836-602-6 (1-301) x US-07-915-246-1 (1-1505)
Qy 12 SerLeuLeuProSerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeu 31
Db 1370 AGTGCATACCTCCCTCTCTCCGCCACCTCCAGCTCCACACACCGCTCTCTCC 1311
Qy 32 ProThrGlnGlyThrPheSer-----AsnValSerThrAsnValSerTyr 46
Db 1310 ACCACCGTGTCTCTCTCTCCACACACACCGATCTCCCTCTCCAGCTCCACCTCCGCGC 1251
Qy 47 GlnGluThrThrProSerThrLeuGlySerThrSerLeuHisProValSerGlnHis 66
Db 1250 ACCATAACACACCGTGTGAGCTCCACCGCGGAGAACCTCTCTCTCCACCGCTCTCTTT 1191
Qy 67 GlyAsnGluAlaThrThrAsnIle-----ThrGluThrThrValLysPheThrSer 83
Db 1190 TCCCCCGCTCCACACCTCTCATGTCCACCTGCACCGCCACCTCCATATCTCTCTCCAGC 1131
Qy 84 ThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSerGlnThrSer 103
Db 1130 ACCAGCGCTCTCCACACCGATTCACACCGTGTGCACCTCCACCTCTCTCCACCGCC 1071
Qy 104 ValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThrThrLeuLys 123
Db 1070 TCCATTTCCACCGCTCTCCACAC-----TCCATGTCCACCTGCACCGCCACAC----- 1023
Qy 124 ProSerLeuSerProGlyAsnValSerAspLeuSerThrThr-----SerThrSerLeuAla 142
Db 1022 ---TCCATATCTCTCTCCACACCGACACCTCTCTCCACACCGATTCACACCATGCGC 966
Qy 143 ThrSerProThrLysProTyrThrSerSerSerProfileLeuSerAspIleLysAlaGlu 162
Db 965 ACCGCCAGCTCTCCACACACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 906
Qy 163 IleLysCysSer 166
Db 905 TCCTCCGCTCTCC 894
RESULT 9
US-09-197-649-7
; Sequence 7, Application US/09197649
; Patent No. 6194550
; GENERAL INFORMATION:
; APPLICANT: Gold, Larry
; APPLICANT: Tuerk, Craig
; APPLICANT: Pribnow, David
; APPLICANT: Smith, Jonathan D.
; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
; FILE REFERENCE: NEX02/C1-CON
; CURRENT APPLICATION NUMBER: US/09/197,649
; CURRENT FILING DATE: 1998-11-23
; EARLIER APPLICATION NUMBER: 07/829,461
; EARLIER FILING DATE: 1992-01-31
; EARLIER APPLICATION NUMBER: 07/739,055
; EARLIER FILING DATE: 1991-08-01
; EARLIER APPLICATION NUMBER: 07/561,968
; EARLIER FILING DATE: 1990-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
; OTHER INFORMATION: fragments having NcoI restriction sites.

US-09-197-649-7

Alignment Scores:		
Pred. No.:	3.6e-05	Length: 390
Score:	138.00	Matches: 32
Percent Similarity:	46.40%	Conservative: 26
Best Local Similarity:	25.60%	Mismatches: 67
Query Match:	4.48%	Indels: 0
DB:	4	Gaps: 0

US-09-836-602-6 (1-301) x US-09-197-649-7 (1-390)

[illegible]

RESULT 10

```

US-08-072-610-1/c
; Sequence 1, Application US/08072610
; Patent No. 5532133
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; and Diagnostic Assays
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,610
; FILING DATE: 19930602
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/07686
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687

```

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Darby and Darby
;; STREET: 805 Third Ave.
;; CITY: New York
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10022-7513
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; FILING DATE: 09/30/96
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Gogoris, Adda
;; REGISTRATION NUMBER: 29,714
;; REFERENCE/DOCKET NUMBER: 5986/17686US2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212)527-7700
;; TELEFAX: (212)753-6237
;; TELEX: 236687
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3337 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORGANISM: Plasmodium vivax
;; IMMEDIATE SOURCE:
;; CLONE: PVM3.3.1
US-08-719-822B-1

Alignment Scores:
Pred. No.: 0.00166 Length: 3337
Score: 127.50 Matches: 53
Percent Similarity: 41.88% Conservative: 27
Best Local Similarity: 27.75% Mismatches: 96
Query Match: 8.45% Indels: 15
Gaps: 6

US-09-836-602-6 (1-301) x US-08-719-822B-1 (1-3337)

Qy 5 TrpThrAlaLeuCysLeuLeuSerLeuLeuPro-----SerGlyPheMetSerLeu 21
Db 2527 TGGTAGTCGATGCGATGATAGATAAAATCCAGAGCCCTAATTCGCTCTTTCTTTGGTTG 2468
Qy 22 AsnAsn-----AsnGlyThrAlaThrProGluLeuProThrGlnGlyThr 36
Db 2467 TCTAATTCATATCTCTTCGCGATGATTAATAATGTAATACATCTCTCTCTCTACT 2408
Qy 37 PheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGly 56
Db 2407 GGTCTTCCT 2348
Qy 57 SerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGlu 76
Db 2347 GGTACTTCTACTTCTACTACCGCTGGTACCTCTACTTCTCTCTCTCTCTCTCTCTCTCT 2288
Qy 77 ThrThrValLysPheThrSerThrSerValIleThrSerValThrGlyAsnThrAsnSer 96
Db 2287 ACTTCCTCTGCTACT 2228
Qy 97 SerValGlnSerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSer 116
Db 2227 TCCACTTCTCTGGTACCTCT-----TCCACTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2174

Qy 117 ThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThr 136
Db 2173 ACTTCTTCTGGTACCTCTTCCACTTCTTACTTCTGAGGTACCTCT-----TCCACT 2120
Qy 137 ThrSerThrLeuAlaThrSerProThr---LysProTyrThrSerSerSerProIle 155
Db 2119 TCTTCTACTTCTGAGGTACCTCTTCCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2063
Qy 156 LeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGln 175
Db 2062 ---TCTACGGTTCTTCTACTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2006
Qy 176 GlyIleCysLeuGluGlnAsnLysThrSerSer 186
Db 2005 GGAGTTGGCTCTAATCTTCTTCTGCGACTTCT 1973
RESULT 12
US-09-092-458-1/c
; Sequence 1, Application US/09092458
; Patent No. 6231861
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,458
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/719,821
; FILING DATE: 09/30/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/17686US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3337 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: PVM3.3.1
US-09-092-458-1

Alignment Scores:
Pred. No.: 0.00166 Length: 3337
Score: 127.50 Matches: 53
Percent Similarity: 41.88% Conservative: 27
Best Local Similarity: 27.75% Mismatches: 96
Query Match: 8.45% Indels: 15

```
DB: 4 Gaps: 6
US-09-836-602-6 (1-301) x US-09-092-458-1 (1-3337)
QY 5 TrpThrAlaLeuCysLeuLeuSerLeuLeuPro-----SerGlyPheMetSerLeu 21
DB 2527 TGGTAGTCATCGATGATGATAAATAATCCAGAGCCCTAAATCGTCTTCTTTGTTG 2468
QY 22 AspAsn-----AsnGlyThrAlaThrProGluLeuProGlnGlyThr 36
DB 2467 TCTAATGTATATCTCTCCGATGGTATTAATAATGTAATACATCTTCTCCCTACT 2408
QY 37 PheSerAsnValSerThrAsnValSerTyGlnGluThrThrProSerThrLeuGly 56
DB 2407 GGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2348
QY 57 SerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGlu 76
DB 2347 GGTACTTCTACTCTACTACCGCTGGTACCTCTACTTCTCTACTCTTCCACTTCTTCC 2288
QY 77 ThrThrValLysPheThrSerThrValIleThrSerValTyGlyAsnThrAsnSer 96
DB 2287 ACTTCTCTGGTACCTCTTCCACTCTCTCTGGTACTCTTCCACTCTTCTGGTACTCT 2228
QY 97 SerValGlnSerGlnThrSerValIleSerThrValPheThrProAlaAsnValSer 116
DB 2227 TCCACTTCTCTGGTACCTCT-----TCCACTTCTCTACTCTCTCGGGGTACCTCTTCT 2174
QY 117 ThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThr 136
DB 2173 ACTTCTCTGGTACCTCTTCCACTCTCTCTACTTCTTCCAGGTACCTCT-----TCCACT 2120
QY 137 ThrSerThrSerLeuAlaThrSerProThr---LysProTyThrSerSerSerProIle 155
DB 2119 TCTTCTACTCTCGAGGTACCTCTTCCACTCTTCTCTCTCTCGCAGTTCTTCCGCC--- 2063
QY 156 LeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGln 175
DB 2062 ---TCTACGGTTCTCTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2006
```

```
RESULT 13
US-08-325-267A-3
; Sequence 3, Application US/08325267A
; Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI
; APPLICANT: TAKATA, YOSHIHIRO
; APPLICANT: OGAWA, MASAHIRO
; APPLICANT: PENTTILA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, Sirkka
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; NUMBER OF INVENTIONS: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,267A
; FILING DATE: 18-NOV-1994
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: JP PCT/JP94/00290
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA: JP 38871/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2589 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; STRAIN: ABXL-ID
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2589
; US-08-325-267A-3
Alignment Scores:
Pred. No.: 0.00177 Length: 2589
Score: 125.50 Matches: 72
Percent Similarity: 44.03% Conservative: 46
Best Local Similarity: 26.87% Mismatches: 90
Query Match: 8.32% Indels: 61
DB: 1 Gaps: 12
US-09-836-602-6 (1-301) x US-08-325-267A-3 (1-2589)
QY 26 ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnValSer 45
DB 1099 ACAACACACACTGAA---CCATGGACTGGTACTTTTACTTCGACTTCCACTGAAATGTCT 1155
QY 46 -----TyrGlnGluThr-----ThrPro 52
DB 1156 ACTGTCACTGGAAACCAATGGCTTGCACACTGATGAACACTGTCACTGTGTCAAAACTCCA 1215
QY 53 SerThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThr 72
DB 1216 ACTACTGCCATCTCATCCAGTTTGCA-----TCATCATCTTCAGGACAAATCACCAGC 1269
QY 73 AsnIleThrGluThrThrValLysPheThrSerThrSerValIleThrSerValTyGly 92
DB 1270 TCTATCAGC-----TCTTCGCGTCCCAATATTATACCCCATTTCTATCT 1311
QY 93 AsnThrAsnSerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrPro 112
DB 1312 AGCAATGGAACTTCTGTGATTTCT---TCCTCAGTAATTTCT----- 1350
QY 113 AlaAsnValSerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSer 132
DB 1351 TCTCAGTCACCTCTCTCTATTCATTCCTTCTCCAGTCATTTCTTCTCAGTCATTTCT 1410
QY 133 AspLeuSerThrThrSerThrSerLeu-----AlaThrSerPro 145
DB 1411 TCTTCTACAAACAACCTCCACTTCTATATTTCTGAATCATCTAAATCATCCGCTCATTTCCA 1470
QY 146 ThrLysProTyThrSerSerSerProIleLeuSerAspIleLysAlaGluIleLysCys 165
DB 1471 ACCAGTAGTTCACCTCTCTGGTTCT-----TCTGAGAGCGAAGACG 1509
QY 166 SerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGluGlnAsnLysThrSer 185
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:10:53 : Search time 1044.68 seconds
(without alignments)
3888.820 Million cell updates/sec

Title: US-09-836-602-6
Perfect score: 1509
Sequence: 1 MPRGTALCLLSLPPSGFMS.....LVTSGALLAVLGITGYFLMN 301

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_p2n.model -DEV=xlp
-Q/cn2_1/USPTO_spool/US09836602/runat_29102002_101144_3812/app_query.fasta_1.1429
-DB=EST -OFMT=fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HICAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602.qcgn.1.1.2453 -runat_29102002_101144_3812 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGESQRY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMECOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_gss.*
13: em_gss_hum.*
14: em_gss_inv.*
15: em_gss_pln.*
16: em_gss_vrt.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result	Query	Score	Match Length	ID
No.	Score	Match	Length	Description
1	1509	100.0	977	9 AL548050
2	1409	93.4	919	9 AL513812

3	1153	76.4	735	9	AL598821
4	1122	74.4	761	9	AU141253
5	1096	72.6	752	9	AU141111
6	1078	71.4	749	9	AU121093
7	1069.5	70.9	768	9	AU141883
8	1040	68.9	1129	10	BI769224
9	994	65.9	804	10	EG675028
10	991.5	65.7	754	9	AU135182
11	907.5	60.1	1001	10	BG256232
12	792	52.5	717	9	AU121843
13	755.5	50.1	733	9	AU141125
14	714	47.3	825	9	AU139201
15	685.5	45.4	903	10	BI824944
16	627	41.6	464	10	BM146840
17	627	41.6	551	10	T80274
18	624	41.4	477	10	BM146845
19	610	40.4	839	10	BF178239
20	609	40.4	780	9	A1806138
21	571	37.8	603	9	AA673625
22	553	36.6	657	9	AL547643
23	541	35.9	382	10	C18678
24	514	34.1	464	9	AA792789
25	496	32.9	473	10	BF830285
26	490	32.5	486	9	A1893233
27	468	31.0	424	10	W65699
28	467	30.9	477	9	A1371138
29	416	27.6	805	10	BG864235
30	411	27.2	806	10	BE913124
31	405	26.8	997	9	AL547847
32	390	25.8	397	10	BF993687
33	365	24.2	223	10	H72215
34	361.5	24.0	675	9	BB665591
35	350	23.2	558	9	AW656254
36	338	22.4	297	10	BF471335
37	313.5	20.8	551	9	AW910263
38	302	20.0	263	9	AW839633
39	297.5	19.7	1081	10	BE916866
40	287.5	19.1	609	9	BB659396
41	281.5	18.7	575	9	AA907722
42	273.5	18.1	653	9	BB642054
43	257	17.0	549	10	BE948643
44	253	16.8	248	9	AA311473
45	249	16.5	474	10	BE847358

ALIGNMENTS

RESULT 1
AL548050
LOCUS AL548050 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI034YP22 5
DEFINITION AL548050 977 bp mRNA linear EST 16-FEB-2001
prime, mRNA sequence.
ACCESSION AL548050
VERSION AL548050.1 GI:12882692
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 977)
AUTHORS Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequenace
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..977
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DI034YP22"
/clone_lib="LTI_NFL006_PL2"

/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 259 a 272 c 231 g 214 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 2 64e-125 Length: 977
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-6 (1-301) x AL548050 (1-977)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 40 ATGCCGGGGCTGGACCGCGCTTTGCTGCTGAGTTTGCTGCTTCTGGGTTTCATGAGT 99
Qy 21 LeuAspAsnGlnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 100 CTTGACAAACACGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAATGTT 159
Qy 41 SerThrAsnValSerTyrglnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 160 TCTCAATGATGCTCTACCAAGAACTACACACTAGTACCTTGGAGTACAGGCTG 219
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 220 CACCCTGTGCTCAACATGCAATGAGCCACACAAACATCACAGAAACAGACAGTCAAA 279
Qy 81 PheThrSerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValGlnSer 100
Db 280 TTCACTACTACCTCTGTGATAACCTCAGTTTATGGAACACAACTCTTCTGTCCAGTCA 339
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 340 CAGACCTCTGTATACACACAGAGTTCCACCACCCACACAGTTTCACTTCCAGAGACA 399
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 400 ACCTTGAAGCTAGCCTGTACCTGGAAATGTTTCAGACCTTTTCAACCACTAGCCTAGC 459
Qy 141 LeuAlaThrSerProThrLysProTyThrSerSerSerProIleLeuSerAspIleLys 160
Db 460 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTTAAAGTGACATCAAG 519
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 520 GCAGAAATCAATGTTTCAGGCATCAGAGAGTGAATGACTCAGGCACTGCGCTGGAG 579
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 580 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAAGACAGAGGGAGGCGCTGGCCCGA 639
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 640 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGCTGGGCCAGGATGCTCCCTGCTC 699
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 700 CTTGCCAGCTGAGGTGAGGCCTCAGTGTCTACTGCTGTGCTTGGCCACAGACAGAA 759
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260

Db 760 ATTTCCAGCAACTCCAACTATGAAAAGCACCAATCTGACCTGAAAAAGCTGGGATC 819
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrsGlnLysThrLeuLeu 280
Db 820 CTGATTTCACTGAGCAAGATGTTGCAAGCCACAGAGCTATTCCCAAAAGACCTGATT 879
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrsPheLeuMet 300
Db 880 GCATGCTACCTCGGGAGCCCTGCTGGCTGTCTTGGGCATCACTGGCTATTTCTCTGATG 939
Qy 301 Asn 301
Db 940 AAT 942

RESULT 2
AL513812 919 bp mRNA linear EST 13-FEB-2001
LOCUS AL513812 LTI_NFL006_PL2 Homo sapiens cDNA clone CLOBA0092F10 5
DEFINITION prime, mRNA sequence.

ACCESSION AL513812
VERSION AL513812.1 GI:12777306
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 919)

AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

Location/Qualifiers

1..919

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CLOBA0092F10"

/tissue_type="placenta"

/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact : Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 246 a 253 c 213 g 198 t 9 others

ORIGIN

Alignment Scores:

Pred. No.: 2 25e-116 Length: 919

Score: 1409.00 Matches: 286

Percent Similarity: 97.61% Conservative: 0

Best Local Similarity: 97.61% Mismatches: 7

Query Match: 93.37% Indels: 1

DB: 9 Gaps: 0

US-09-836-602-6 (1-301) x AL513812 (1-919)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20

Db 40 ATGCCGGGGCTGGACCGCGCTTTGCTGCTGAGTTTGCTGCTTCTGGGTTTCATGAGT 99

Qy 21 LeuAspAsnGlnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40

Db 100 CTTGACAAACACGGTACTGCKACCCAGAGTTTACCTACCAGGGGAACATTTTCAATGTT 159


```
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 160 TCTACAAATGTATCCYACCAAGAAATACACACACCTAGTACCTTGGAGTACCGCTG 219
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 220 CACCCTGTGTCTCAACATGGCAATGAGGCCACACAAACATCACAGAACGACAGTAAA 279
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 280 TTCACATCTACCTCTGTATACCTCAGTTTATGGAACACAAACATCTCTCTCCAGTCA 339
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 340 CAGACCTCTGTATATCAGCAGACAGTGTACACACCCAGCCCAACGTTTCAACTCCAGAGACA 399
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 400 ACCTTGAAGCCTAGCTGTACCTGGAAATGTTTTCAGACCTTTCAACCACTAGCAGCTAGC 459
QY 141 LeuAlaThrSerProThrLysProThrLysProThrSerSerProIleLeuSerAspIleLys 160
Db 460 CTTCGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCCYAAAGTACATCAG 519
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 520 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTCAGCTCAGGGCATCTCCCTGGAG 579
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlyLeuAlaArg 200
Db 580 CAAAATAAGACCTCCACGCTGTCGGAGGTTTAAAGAGGACAGGGAGGGCCCTGGCCCA 639
QY 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 640 GTGCTGTGTGGGAGGAGGAGGCTGATGCTGATGCTGGGGCCCGGATGCTCCCTGCTC 699
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 700 CTTCGCCAGTCTGAGGTGAGGCTCAGTGTCTACTGCTGTCTGGCCCAACAGACAGAA 759
QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 760 ATTTCAGCAAACTCCAACTTATGAAAAGACCAATCTGACCTGAAAAGCTGGGGATC 819
QY 261 LeuAspPheThrGluGlnAspValAlaLaserHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 820 YTATGATTTCACTGAGCMAGAGTGTGCAACCCAGCAGATATTCCTCCAAAGACCTGATT 879
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGly 293
Db 880 GCACTGGTCACCTCGGGAGC-CTGCTGGSTGTCTTGGGC 917

RESULT 3
AL598821 735 bp mRNA linear EST 14-AUG-2001
DEFINITION DKF2p313K2322_r1 313 (synonym: hlcc2) Homo sapiens cDNA clone
ACCESSION AL598821
VERSION AL598821.1 GI:15161512
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 735)
AUTHORS Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann
,S.
TITLE EST (Duesterhoeft, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: Duesterhoeft A
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
```

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.

No sl sequence available.
This clone (DKF2p313K2322) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

Location/Qualifiers

1..735

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="DKF2p313K2322"

/clone_lib="313 (synonym: hlcc2)"

/dev_stage="adult"

/lab_host="DH10B"

/note="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiIb;

cDNA-collection"

BASE COUNT 191 a 205 c 173 g 164 t 2 others

ORIGIN

Alignment Scores:

Pred. No.: 1.5e-93 Length: 735

Score: 1153.00 Matches: 233

Percent Similarity: 98.31% Conservative: 0

Best Local Similarity: 98.31% Mismatches: 2

Query Match: 76.41% Indels: 2

DB: 9 Gaps: 0

US-09-836-602-6 (1-301) x AL598821 (1-735)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuProSerGlyPheMetSer 20

Db 28 ATGCCCGGGGCTGGACCGCGCTTGTCTGCTGAGTTTCTGCCTCTGGGTTCATGAT 87

QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40

Db 88 CTTGACAACAACGGTACTGCTACCCGAGAGTTACCTACCCAGGAGCAATTTCAATGTT 147

QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60

Db 148 TCTCAAAATGTATCTTACCAAGAACTACACACCTTAGTACCTTGGAGTAGTACCGCTG 207

QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80

Db 208 CACCCTGTCTCAACATGCAATGAGGCCACACAAACATCACAGAAACAGACAGTCAA 267

QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100

Db 268 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACCTTCTGTCTCCAGTCA 327

QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120

Db 328 CAGACCTCTGTATCAGCACAGTGTTCACACCCCGCCAGCCAGTTCCTCACTCCAGAGACA 387

QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140

Db 388 ACCTTGAAGCCTAGCTGTACCTGGAAATGTTTTCAGACCTTTCACCACTAGCAGTACG 447

QY 141 LeuAlaThrSerProThrLysProThrLysProThrSerSerProIleLeuSerAspIleLys 160

Db 448 CTTGCAACATCTCCCACTAAACCTTATACATCATCTTCTCTATCTTCAAGTACATCAAG 507

QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180

Db 508 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTCAGCTCAGGGCATCTGCTGGAG 567

QY 181 GlnAsnLysThrSerSerCysAlaGluPheLys-LysAspArg-GlyGluGlyLeuAla 200

Db 568 CAAAATAAGACCTCCAGCTGTGCGGAGTTTAAAGAGGACAGGGGAGAGGCCCTGGCCC 627

QY 200 rgValLeuCysGlyGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuL 220

|||||

```
Db 628 GAGTGTCTGTGGGAGGAGCAGCTGATGCTGATGCTGGGGCCCGCAGGTATGCTNCCTGC 687
Qy 220 euLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeu 235
|||||
Db 688 TCCCTGGCCAGCTGAGGNGAGGCTCAGTGTCTACTGCTGCTTG 734
|||||

RESULT 4
LOCUS AU141253 761 bp mRNA linear EST 25-OCT-2000
DEFINITION AU141253 THYR01 Homo sapiens cDNA clone THYR01000286 5', mRNA
sequence.
ACCESSION AU141253
VERSION AU141253.1 GI:11002774
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 761)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source Location/Qualifiers
1..761
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01000286"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/note="Vector: pME18SFL3"
BASE COUNT 197 a 214 c 180 g 167 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 9.51e-91 Length: 761
Score: 1122.00 Matches: 222
Percent Similarity: 98.68% Conservative: 2
Best Local Similarity: 97.80% Mismatches: 3
Query Match: 74.35% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-6 (1-301) x AU141253 (1-761)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 73 ATGCGCGGGGCGTGGACCGCGTTGCTGCTGAGTTGCTGCTTCTGGTTTCATGAGT 132
|||||
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 133 CTTGACAAACACGGTACTGTACCCAGAGTTACCTACCCAGGGAACATTTCAATGTT 192
|||||
Qy 41 SerThrAsnValSerTyrglnGluThrThrThrProSerThrLeuGlySerThrLeu 60
|||||
Db 193 TCTACAAATGTATCTTACCAAGAAACTACAAACACCTAGTACCTTGGAAAGTACCAGCTG 252
|||||
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
|||||
Db 253 CACCCTGTGCTCAACATGCAATGAGGCCACACAGCATCACAGAAACGACAGTCAA 312
|||||
Qy 81 PheThrSerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValGlnSer 100
|||||
```

```
Db 313 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 372
|||||
Qy 101 GlnThrSerValIleSerThrValPheThrThrProIlaasnValSerThrProGluThr 120
|||||
Db 373 CAGACCTCTGTAATCAGCAGAGTGTTCACCACCCACCAACGTTTCAACTCCAGAGACA 432
|||||
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
|||||
Db 433 ACCTTGAAGCTAGCCCTGTCACCTGGAAATGTTTCACACCTTTCACACCTAGCATTAGC 492
|||||
Qy 141 LeuAlaThrSerProThrLysProTyThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 493 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTTCTTCTTCTTCTTCT 552
|||||
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
|||||
Db 553 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAAGTGAATTTGACTCAGGGCATCTCGCTGGAG 612
|||||
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
|||||
Db 613 CAAAATAGACCTTCCAGCTGTGGGAGTTTAAAGAGACAGGGGAGAGGGGCTGGCCCGA 672
|||||
Qy 201 ValLeuCysGlyGluGlnGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
|||||
Db 673 GTGCTGTGTGGGAGGAGCANGCTGATGCTGATGCTGGGCCCAAGTATGCTCCCTGNTC 732
|||||
Qy 221 LeuAlaGlnSerGluValArg 227
|||||
Db 733 CTTGCCCACTCTGANGTGAAG 753
|||||

RESULT 5
LOCUS AU141111 752 bp mRNA linear EST 25-OCT-2000
DEFINITION AU141111 THYR01 Homo sapiens cDNA clone THYR01000038 5', mRNA
sequence.
ACCESSION AU141111
VERSION AU141111.1 GI:11002632
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 752)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source Location/Qualifiers
1..752
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01000038"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/note="Vector: pME18SFL3"
BASE COUNT 197 a 208 c 176 g 168 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 2.01e-88 Length: 752
```

Score: 1096.00 Matches: 218
Percent Similarity: 98.20% Conservativeness: 0
Best Local Similarity: 98.20% Mismatches: 4
Query Match: 72.63% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-6 (1-301) x AU141111 (1-752)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 65 ATGCCCGGGGCTGACCGCGCTTGGCTGCGAGTTGCTGCCCTCTGGGTTCATGAGT 124
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 125 CTTGACAACACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAATGTT 184
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 185 TCTACAAATGTTATCTTACCAAGAACTACACACCTAGTACCTTGGAAAGTACCGACCTG 244
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
|||||
Db 245 CACCCTGTGCTCAACATGGCAATGAGGCCACACAAACATCACAGAAACGACAGTCAA 304
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
Db 305 TTTACATCTTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 364
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||||
Db 365 CAGACCTCTGTAATCAGCACAGTGTACACCCCGCCAGCAAGTTCACACTCCAGAGACA 424
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 425 ACCTTGAAGCCTAGCTGTACCTGGAATGTTTACAGACCTTTCAACCACTAGCAGTACG 484
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 485 CTTGCAACATCTCCACATAACCCCTATACATCATCTTCTCTATCTTAAAGTGACATCAAG 544
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
|||||
Db 545 GCAGAAATCAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCAGGGCATCTGCCTGGAG 604
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysAspArgGlyGluGlyLeuAlaArg 200
|||||
Db 605 CAAAATAAGACCTTCAGCTGTGCGAGTTTAAAGAGGAGGGGAGGGGCTGGCCCGA 664
Qy 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
|||||
Db 665 GTGCTGTGGGGAGGAGCANGCTGATGCTGATGCTGGGGCCCAAGTATTGCTCTGCTC 724
Qy 221 LeuAla 222
|||||
Db 725 CTTGCC 730

RESULT 6
AU121093
LOCUS AU121093 HEMBB1 Homo sapiens cDNA clone HEMBB1002065 5', mRNA
DEFINITION
VERSION AU121093.1 GI:10936328
KEYWORDS EST.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 749)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project

JOURNAL
COMMENT

Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5', & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES
source

Location/Qualifiers
1..749
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBB1002065"
/clone_lib="HEMBB1"
/tissue_type="whole embryo, mainly body"
/dev_stage="embryo, 10 weeks"
/note="Vector: pME18SFL3"
BASE COUNT 195 a 211 c 173 g 166 t 4 others
ORIGIN

Alignment Scores:

Pred. No.: 8.22e-87 Length: 749
Score: 1078.00 Matches: 214
Percent Similarity: 96.83% Conservativeness: 0
Best Local Similarity: 96.83% Mismatches: 7
Query Match: 71.44% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-6 (1-301) x AU121093 (1-749)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 73 ATGCCCGGGGCTGACCGCGCTTGGCTGCGAGTTGCTGCCCTCTGGGTTCATGAGT 132
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 133 CTTGACAACACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAATGTT 192
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 193 TCTACAAATGATCTCTTACCAAGAACTACACACCTAGTACCTTGAAGTACCGCTG 252
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
|||||
Db 253 CACCCTGTGCTCAACATGGCAATGAGGCCACAAACATCACAGAAACGACAGTCAA 312
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
Db 313 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 372
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||||
Db 373 CAGACCTCTGTAATCAGCACAGTGTTCACACCCCGCCAGCAACCTTCAACTCCAGACA 432
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 433 ACCTTGAAGCCTAGCCCGCTCACCTGGAAATGTTTTCAGACCTTTTCAACCACTAGCAGTACG 492
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 493 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTTAAAGTACATCAAG 552
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
|||||
Db 553 GCAGAAATCAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCANGGCATCTGCCTGGAG 612
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
|||||
Db 613 CAAAATAAGACCTTCAGCTGTGCGAGTTTAAAGAGGAGGGGAGGGGCTGGCCCGA 672

```

QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
|||||
Db 673 GTGCTGTGGGGAGAACANGCTGATGCTGATGCTGGGGCCCAAGTATGCTCTGTCTCC 732
QY 221 Leu 221
|||
Db 733 TTG 735

RESULT 7
AUI141883 768 bp mRNA linear EST 25-SEP-2000
LOCUS AUI141883 THYR01 Homo sapiens cDNA clone THYR01001399 5', mRNA
DEFINITION
ACCESSION AUI141883
VERSION AUI141883.1 GI:11003404
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 768)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamanoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1..768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01001399"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/notes="Vector: pME18SFL3"
BASE COUNT 198 a 215 c 181 g 171 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 4,93e-86 Length: 768
Score: 1069.50 Matches: 217
Percent Similarity: 95.61% Conservative: 1
Best Local Similarity: 95.18% Mismatches: 10
Query Match: 70.87% Indels: 1
DB: 9 Gaps: 0

US-09-836-602-6 (1-301) x AUI141883 (1-768)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 72 ATGCCGGGGGCTGGACCGGCTTTGCTGCTGAGTTTGCTGCTTCTGGGTTTCATGAGT 131
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 132 CTTTGACAAACACGGTACTGTACCCAGAGTTACCTACCCAGGGGAACATTTCAAATGTT 191
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 192 TCTACAATGTATCTACCAAGAACTACAACTAGTACCTTGGAGTACAGGCTG 251
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
|||||

```

```

Db 252 CACCTGTGTCTCAACATGGCAATGAGCCACAAACACACAGAAACACAGAGTCAAA 311
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
Db 312 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAAACCTCTCTGTCCAGTCA 371
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||||
Db 372 CAGACCTCTGTAATCAGCACAGTGTTCACACCCAGCCAGCGTTTCAACTCCAGAGACA 431
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 432 ACCTTGAAGCCTAGCCTGCTCACTACCTGGAAATGTTTCAGACCTTTCAACCACTAGCACTAGC 491
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 492 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTTCAAGTGACATCAAG 551
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
|||||
Db 552 GCAGAAATCAATGTTTCAGGCATCAGAGAGTGAATTGACTCAGGGCATCTGCCTGGAG 611
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
|||||
Db 612 CAAATAAGACCTTTCAGCTGTGCGAGTTTAAAGAAGACAGAGGGAGGCGCTGCCCGCA 671
QY 201 ValLeuCysGlyGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
|||||
Db 672 GTGCTGTGGGGAGGAGAACANGCTGATGCTGATGCTGGC--CCAGTATGCTTCTCTGCTC 729
QY 221 LeuAlaGlnSerGluValArgPro 228
|||||
Db 730 CTTGCCAATCTGANGTGAAGGCCT 753

RESULT 8
BI769224 1129 bp mRNA linear EST 25-SEP-2001
LOCUS 603060051f1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5209333 5',
DEFINITION mRNA sequence.
ACCESSION BI769224
VERSION BI769224.1 GI:15760802
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1129)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@email.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11525 row: e column: 14
High quality sequence stop: 765.
FEATURES
source
1..1129
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5209333"
/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/notes="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed)

```

upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC Library.*

BASE COUNT 292 a 317 c 279 g 241 t
ORIGIN

Alignment Scores:

Pred. No.: 3,79e-83 Length: 1129
Score: 1040.00 Matches: 247
Percent Similarity: 81.70% Conservativity: 12
Best Local Similarity: 77.92% Mismatches: 31
Query Match: 68.92% Indels: 27
DB: 10 Gaps: 7

US-09-836-602-6 (1-301) x B1769224 (1-1129)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 55 ATGCCGCGGGCTGGACCGCGCTTGGCTGCTGAGTTGCTGCCTTCTGGGTTCATGAGT 114
Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 115 CTTGACAAACAGGCTACTGCTACCCAGAGTACTACCCAGGGAAACATTTTCAAATGTT 174
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 175 TCTACAAATGATCTCTACCAAGAACTACACACCTAGTACCTTGGAGTACAGCCCTG 234
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 235 CACCCTGTGTCTCAACATGGCAATGAGGCCACACAAACATCACAGAAACGACAGTCAA 294
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 295 TTCACATCTACCTCTGTGATACCTCAGTTTATGGAAACACAAACTCTCTGTCCAGTCA 354
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 355 CAGACCTCTGTANYCAGCACAGTGTTCACACCACCCGCGCAAGTTCACCTCCAGAGACA 414
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 415 ACCTTGAAGCCTAGCCTGCTACCTGGAATGTTTTCAGACCTTTCACACCTAGCAGTAC 474
Qy 141 LeuAlaThrSerProThr-LysProTyrThrSerSerSerProIleLeuSerAspIlely 160
Db 475 CTTGCAACATCTCCCACTAAAGCCCTATACATCATCTCTCTATCTCTAAGTGACATCAA 534
Qy 160 sAlaGluIleLysCysSerGlyIle-ArgGluVal-LysLeuThrGlnGlyIleCysLeu 179
Db 535 GGCAGAAATCAATGTTTCAGGCATCACAGAAAGTGACAAATGACATCAGGCGATCGCCTG 594
Qy 180 GluGlnAsnLysThrSerSerCysAlaGluPheLysAspArgGlyGluGlyLeuAla 199
Db 595 GAGCAAAATAAGACCTCCAGCTGTGCGAGTTTAAAGAGGACAGGGAGGGCCCTGGCC 654
Qy 200 -ArgValLeuCys---GlyGluGluGluAlaAspAlaGlyAla-Gln---ValC 217
Db 655 CCGAGTGTGTGCTGGCGGAGGACAGCGCTGATGCTGATGCTGGGGCCCGACGCTATGC 714
Qy 217 ySerLeuLeuAlaGlnSerGlu-ValArgPro-GlnCysLeuLeuValLeuAl 236
Db 715 TCCCGGTGGTCTTCCCAAGCTGAGCGTGAGGCCCTCCAGTGTCTACTGCTGGGTCTTGC 774
Qy 236 a-----AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisG 252
Db 775 GCGGCTAGGAAGCGGAAATTTGCCCGACCAAGCGTCCCAAGCTTTATGAAAAGCACCCA 834
Qy 252 nSerAspLeuLysLysLeu-----GlyIleLeuAspPheThrGluGlnAspValAlaSe 270
Db 835 ATCTGAGCCTGAAACAGAGCTGGGGGAATCCATAGAATTTTCGCTGAGCAAGATTTGTCGAC 894

Qy 270 rHisGlnSerTyrSerGlnLysThrLeuIleAla-----Le 282
Db 895 AAGGCATAGC-----AGAAACTTATTTCCTCCAGAGAGCCCTGATTGGCACCT 945
Qy 282 uValThr---SerGlyAlaLeuLeuAlaValLeuGly 293
Db 946 GGAGACCCCTCGGGAAGCCCTGCTGGATGTTACTTGGG 982
RESULT 9
BG675028 804 bp mRNA linear EST 01-MAY-2001
LOCUS 602621258F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4746591 5',
DEFINITION mRNA sequence.
ACCESSION BG675028
VERSION BG675028.1 GI:13906424
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 804)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10595 row: d column: 16
High quality sequence stop: 686.
Location/Qualifiers
FEATURES
source
1..804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4746591"
/clone_lib="NCI_CGAP_Skn3"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. primer: Oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 219 a 217 c 191 g 177 t
ORIGIN
Alignment Scores:
Pred. No.: 3,11e-79 Length: 804
Score: 994.00 Matches: 229
Percent Similarity: 91.41% Conservativity: 5
Best Local Similarity: 89.45% Mismatches: 10
Query Match: 65.87% Indels: 12
DB: 10 Gaps: 2
US-09-836-602-6 (1-301) x BG675028 (1-804)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 46 ATGCCGCGGGCTGGACCGCGCTTGGCTGCTGAGTTGCTGCCTTCTGGGTTCATGAGT 105
Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 106 CTTGACAAACAGGCTACTGCTACCCAGAGTACCTACCCAGGGAACATTTTCAAATGTT 165
Qy 41 SerThr-AsnValSerTyrGlnGlu---ThrThrThrProSerThrLeuGlySerThrSe 59
Db 166 TCTCAAAAGTGTATCTCTACCAAGAAAGTACACAGCCTAGGTACCTTGTGTTAAGTACCAG 225
Qy 59 rLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVa 79

```
|||||
Db 226 CCTGCACCTGTGTCACACATGGCAATGAGGCCACAAACAAACATCACAGAAACGACAGT 285
Qy 79 llySerPheThr-SerThrSerValIleThrSerValThrGlyAsnThrAsnSerSerValG 99
Db 286 CAARTTCAGCATCTACTCTGATGATACCTCAGTTTATGAAACAAACACTCTCTGTCC 345
Qy 99 lnSerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProG 119
Db 346 AGTCACAGACCTCTGTAATCAGCACAGTGTTCACCAACCCAGCAAGCTTTTCAACTCCAG 405
Qy 119 luThrThr-LeuLysProSerLeuSerProGlyAsnValSerAsp-LeuSerThrThr-S 138
Db 406 AGACAAGCCTTGAAGCCTAGCCTGTACCTGGAAATGTTTCAGACCCCTTTCAACCACTAG 465
Qy 138 erThrSerLeuAlaThrSerProThrLysProThrLysProThrLysSerSerProIleLeuSerA 158
Db 466 GCNCTAGCCTTGCACATCTCCCACTTAACCCATATACATCATCTTCTCTATCCTTAAGTG 525
Qy 158 spileLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleC 178
Db 526 ACATCAAGCAGAAATCAATGTTTCAGGCATCAGAGAGTGAATTTGACTCAGGGCATCT 585
Qy 178 ysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyL 198
Db 586 GCCTGGAGCAAAATAAGACTTCCAGCTGTGCGGAGTTTAAGAAGAGACAGGGGAGGGAC 645
Qy 198 euAlaArgValLysCysGlyGluGluGln-AlaAspAlaAsp-AlaGlyAlaGlnVal-- 216
Db 646 TGCCCGAGTGTGTGTGGGGAGGAGCAGAGCTGATGCTGATAGCTGGGGGCCAGGTATA 705
Qy 217 -CysSerLeuLeuAlaGlnSer-GluValArgProGlnCysLeuLeu-LeuValLeu 235
Db 706 GCTCCCTGCTACCTTGCCCGCAGCTAGAGGTGAGGCTCAGTCTACTGACTGGTCCCTG 765
Qy 236 -AlaAsnArgThrGluIleSerSerLys 244
Db 766 GGCAAAACAGAACAGACATTTTCAGCAAAA 793

RESULT 10
AUI35182 754 bp mRNA linear EST 24-OCT-2000
LOCUS AUI35182 PLACE1 Homo sapiens cDNA clone PLACE1001404 5', mRNA
DEFINITION sequence.
ACCESSION AUI35182
VERSION AUI35182.1 GI:10995721
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 754)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Location/Qualifiers
1..754
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1001404"
source
```

```
/clone_lib="PLACE1"
/tissue_type="placenta"
/note="vector: pME18SPL3"
BASE COUNT 201 a 210 c 172 g 166 t 5 others
ORIGIN
Alignment Scores:
Pred. No.: 4,76e-79 Length: 754
Score: 991.50 Matches: 208
Percent Similarity: 90.60% Conservative: 4
Best Local Similarity: 88.89% Mismatches: 12
Query Match: 65.71% Indels: 10
DB: 9 Gaps: 3
US-09-836-602-6 (1-301) x AUI35182 (1-754)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 48 ATCCCGCGGGGCTGGACCGCGCTTGTGCTGAGTTGCTGCTTCTGGGTTTCATGAGT 107
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 108 CTTGACAAACACCGTACTGCTACCCAGAGTTACTACCCAGGGGAACATTTTCAATGTT 167
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 168 TCTACAAATGTATCCTTACCAAGAAACTACAACACCTAGTACCTTGGAGTACCAGCCTG 227
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 228 CACCCCTGTCTCAACATGGCAATGAGCCACCAACAAATCATCACAGGACAGTCAAA 287
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 288 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAAACACAAACTCTTCTGTCCAGTCA 347
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 348 CAGACCTCTGTAATCAGCACAGTGTTCACCAACCCAGCCAACTTTCAACTCCAGAGACA 407
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerSer 140
Db 408 ACCTTGAAGCCTACCTGTGCCTGGAAATGTTTCAGACCTTTCAACCACTAGCCTAGC 467
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 468 CTTGCAACATCTCCCACTAAACCTTATACATCATCTTCTCTATCTTAAGTGAACATCAAG 527
Qy 161 Ala-GluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuG1 180
Db 528 GCAAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAAATTCACCTCAGGGCATCTGCCTGGA 587
Qy 180 uGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu---GlyLeuAl 199
Db 588 GCAAAATAAGACCTCCAGCTGTGCGGAGTTTAAGAAGGACAGGAGGAAAGGCGCTGGCC 647
Qy 199 aArgVal---LeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSe 218
Db 648 CCAGATTGCTGTGTTGGGGAGGAGCAAGGCTGATGCC-----TG 686
Qy 218 rLeuLeuLeuAlaGlnSerGluValArgProGlnCysLeu 231
Db 687 ATTGCTGGGGGCCCAAGTTANNNGCTCCCTCCCTGCTGCTCCCTT 726
RESULT 11
BG256232 1001 bp mRNA linear EST 13-FEB-2001
LOCUS BG256232
DEFINITION 602369196F1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4477318 5',
mRNA sequence.
ACCESSION BG256232
VERSION BG256232.1 GI:12766048
KEYWORDS EST.
SOURCE human.
```


QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 274 CTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAATGTT 333
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 334 TCTACAAATGTATCCTTACCAAGAAATACAAACACCTAGTACCCCTTGGAAAGTACACGCTG 393
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
|||||
Db 394 CACCCTGTCTCTCAACATGGCAATGAGGCCACAAACAACATCACAGAAGCAGAGTCAAA 453
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
Db 454 TTCACATCTACCTCTCTGATACCTCAGTTTATGGAACACACAACTCTTCTGTGTCAGTCA 513
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||||
Db 514 CAGACCTCTGTATATCAGCACAGTGTTCACACCCAGCCCAACGTTTCAAACCTCCAGAGACA 573
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 574 ACCTTGAACCTAGCCTGTACCTCGAAATGTTTCAGACCTTTCAACCCACTAGCACTAGC 633
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 634 CTTGCAACATCTCCACTAAACCCATATCA-TCATCTTCNTCTATCTTAAGTGACATCAAG 692
QY 161 AlaGluIleLysCysSerGlyIle 168
|||||
Db 693 GCAGAAATCAAATGTTTCANGCATC 716
RESULT 13
AUI141125 733 bp mRNA linear EST 25-OCT-2000
LOCUS AUI141125 THYR01 Homo sapiens cDNA clone THYR01000063 5', mRNA
DEFINITION sequence.
ACCESSION AUI141125
VERSION AUI141125
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 733)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. 733
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01000063"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/note="Vector: pME18SFL3"
BASE COUNT 199 a 212 c 168 g 150 t 4 others
ORIGIN

Alignment Scores: 6.73e-58 Length: 733
Pred. No.: 755.50 Matches: 157
Score: 94.61% Conservative: 1
Percent Similarity: 94.01% Mismatches: 8
Best Local Similarity: 50.07% Indels: 2
Query Match: 9 Gaps: 1
DB: 1
US-09-836-602-6 (1-301) x AUI141125 (1-733)
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 233 ATCCCGCGGGCTGGACCGCGCTTTCCTGCTGAGTTTGCCTTCTGGGTTTCATGAGT 292
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 293 CTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAATGTT 352
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 353 TCTACAAATGTATCCTTACCAAGAAACTACAACACCTAGTACCCCTTGGAAAGTACCGCTG 412
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
|||||
Db 413 CACCCTGTCTCTCAACATGGCAATGAGGCCACAAACAACATCACAGAAGCAGAGTCAAA 472
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
Db 473 TTCACATCTACCTCTCTGATACCTCAGTTTATGGAACACACAACTCTTCTGTGTCAGTCA 532
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||||
Db 533 CAGACCTCTGTATATCAGCACAGTGTTCACACCCAGCCCAACGTTTCAACTCCAGAGACA 592
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 593 ACCTTGAACCTAGCCTGTACCTCGAAATGTTTCAGACCTTTCAACCCACTAGCACTACCT 652
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 653 TGCACATCTCC---ACTAAACCCCTATCA-TCATCTTCNTCTATCTTAAGTGACATCAAG 708
QY 161 AlaGluIleLysCysSerGly 167
|||||
Db 709 GCAGAAATCAAATGTTTCAGGC 729
RESULT 14
AUI139201 825 bp mRNA linear EST 25-OCT-2000
LOCUS AUI139201 PLACE1 Homo sapiens cDNA clone PLACE1010139 5', mRNA
DEFINITION sequence.
ACCESSION AUI139201
VERSION AUI139201
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 825)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of

Db 579 GACATCAAGGCCGAAACTCAAATGTCGGCAATCAGACAAAGCGAAACTGACTCCGCG 638
 Qy 175 GlnGlyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGlu-PheLysLysAspAr 194
 Db 639 CCATCGGCCAAGCAGCAGCAAACTTAAGACCTCCAGACTGGCGGAAGCACACAGAGGACAA 698
 Qy 194 gGlyGluGly 197
 Db 699 GGGCACAGGG 708

Search completed: October 30, 2002, 10:48:00
Job time : 1048.68 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:08:08 ; Search time 1543.01 Seconds
(without alignments)
4082.221 Million cell updates/sec

Title: US-09-836-602-6
Perfect score: 1509
Sequence: 1 MPRGWTALCLLLPSGFMS.....LVTSGALLVLGITGYFLMN 301

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPRO-SPOOL/US09836602/runat_29102002_101144_3785/app_query.fasta_1.1429
-DB=GenEmbl -QPART=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602_CGN_1_1_3800_@runat_29102002_101144_3785 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAB -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=10 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_be:*
2: gb_hcg:*
3: gb_in:*
4: gb_on:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_on:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1509	100.0	2463	9 S53910	S53910 CD34-glycop
2	1509	100.0	2615	9 HUMCD34HS	M81104 Human CD34
3	1509	100.0	2657	6 AX333753	AX333753 Sequence
4	1509	100.0	2657	9 S53911	S53911 CD34-glycop
5	899	59.6	2956	4 CFU49457	U49457 Canis famil
6	850	56.3	2690	4 AF461503	AF461503 Sus scrof
7	839	55.6	1065	6 AR078454	AR078454 Sequence
8	805.5	53.4	978	10 S69301	S69301 CD34-cell s
9	805.5	53.4	1149	10 S69299	S69299 CD34-cell s
10	805.5	53.4	1260	10 S69293	S69293 CD34-stem c
11	805.5	53.4	2398	10 BC006607	BC006607 Mus muscu
12	728	48.2	581	9 AF202879	AF202879 Homo sapi
13	707.5	46.9	2453	4 AB021662	AB021662 Bos tauru
14	681.5	45.2	2420	4 AF259378	AF259378 Capra hir
15	617	40.9	160771	9 HS8812	AL035091 Human DNA
16	617	40.9	182742	2 AC084393	AC084393 Homo sapi
17	617	40.9	185162	2 AL365178	AL365178 Homo sapi
18	617	40.9	212535	2 AL356275	AL356275 Homo sapi
19	409	27.1	254	9 HUMCD34S3	M81940 Human CD34
20	313.5	20.8	98347	2 AL513203	AL513203 Mus muscu
21	313.5	20.8	166333	2 AL513470	AL513470 Mus muscu
22	311	20.6	183	9 HUMCD34S2	M81939 Human CD34
23	260	17.2	157	9 HUMCD34S5	M81942 Human CD34
24	233	15.4	235	6 AX247364	AX247364 Sequence
25	214	14.2	165	9 HUMCD34S7	M81944 Human CD34
26	172.5	11.4	3124	5 MMTHRAD1	Y13978 Gallus gall
27	167	11.1	80272	5 AL591180	AL591180 zebrafish
28	166	11.0	89232	5 AL591175	AL591175 zebrafish
29	162.5	10.8	50569	8 YSCH9196	U11583 Saccharomyc
30	162.5	10.8	184427	14 EHVD20824	U20824 Equine herp
31	160.5	10.6	253273	2 AL669823	AL669823 Mus muscu
32	158.5	10.5	183358	9 AC019041	AC019041 Homo sapi
33	156.5	10.4	2161	5 XELFIMC1X	L02115 Frog integu
34	155.5	10.3	167390	9 AC007263	AC007263 Homo sapi
35	155.5	10.3	182534	10 AC074046	AC074046 Mus muscu
36	154	10.2	3053	8 SCYNR0444W	Z11659 S.cerevisia
37	154	10.2	3137	8 YSCAAGLCS	M60590 S.cerevisia
38	153.5	10.2	167254	9 CNS05TDS	AL357093 Human chr
39	153.5	10.2	172789	2 AC104020	AC104020 Homo sapi
40	153	10.1	165227	2 AC098014	AC098014 Rattus no
41	153	10.1	185337	2 AC095762	AC095762 Rattus no
42	153	10.1	286060	2 AL589862	AL589862 Homo sapi
43	152.5	10.1	196553	2 AC073946	AC073946 Mus muscu
44	152	10.1	12412	3 CELH43E16	AC006669 Caenorhab
45	152	10.1	131274	2 AC096869	AC096869 Rattus no

ALIGNMENTS

RESULT 1

S53910 S53910 2463 bp mRNA linear PRI 08-MAY-1993
LOCUS CD34-glycoprotein expressed in lymphohematopoietic progenitor cells
DEFINITION (alternatively spliced) [human, UT7 cells, mRNA, 2463 nt].
ACCESSION S53910
VERSION S53910.1 GI:264766
KEYWORDS human UT7 cells.
SOURCE

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2463)
AUTHORS Nakamura,Y., Komano,H. and Nakauchi,H.
TITLE Two alternative forms of cDNA encoding CD34
JOURNAL Exp. Hematol. 21 (2), 236-242 (1993)
MEDLINE 93146100
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI glibsq 124007] from the original journal article.
This sequence comes from Fig. 1A.

FEATURES
source
1..2463
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene
1..2463
/gene="CD34"
CDS
91..1248
/gene="CD34"
/note="glycoprotein expressed in lymphohematopoietic
progenitor cells; This sequence comes from Fig. 1A"
/codon_start=1
/product="CD34"
/protein_id="AAB25222.1"
/db_xref="GI:264767"
/translation="MLVRRGARGPRMGWALCLLLPSGFMSLDNGTATPELP
TQTFNSVNTSVSYQFTTPSTLIGSTSLHPVSOHQNEATNTITETTKFTSTSVIISV
YGNATSSVQSQTSVIETTPANVSTPETTLKPSLSPGNVSDLTSTISLSTSPKP
YSVSPILSDIKAEKSCGIREVKLFGQIKFQNKTSSCAEFKDGRGLARVLGCEE
QADADAGAKLLAQSEVRPQCLLVLANRTEISSKLQMKKHQSDLKRLGLDFT
EDVASHQSYQKTLIALTVSGALLAVLIGITGYFLMNRSSWPTGERLGEDPYITENG
GGQGYSSGPGTSPAQGKASVNRGSGENGTCQATSRNGHSARQHVAVDEL"
BASE COUNT 589 a 709 c 580 g 585 t
ORIGIN

Alignment Scores:
Pred. No.: 9.67e-92 Length: 2463
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-836-602-6 (1-301) x S53910 (1-2463)

QY 1 MetProArgGlyThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 127 ATGCCGGGGCTGGAGCCGCGCTTGTCTGCTGAGTTGCTGCTCTCTGGGTTTCATGAGT 186
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 187 CTGTGACACACACGGTACTGCTACCCAGAGTTACCTACCCAGGAGACATTTTCAATGTT 246
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 247 TCTACAAATGTATCTTACCAAGAAACTACAAACCTAGTACCTTGGAAAGTACCAAGCTG 306
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
Db 307 CACCCCTGTGCTCAACATGCGCATGAGGCGCACAAACATCACAGAAACGACAGTCAAA 366
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 367 TTCACATCTACCTCTGTGATACCTCAGTTTATGGAACACAACTCTTCTGTGCTCAGTCA 426
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 427 CAGACCTCTGTAATCAGCAGAGTGTTCACCACCCAGCAACCTTTCAACTCCAGAGACA 486
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrThrSerThr 140
Db 487 ACCTTAAGCCTAGCCTGTACCTCGGAAATGTTTTCAGACCTTTCAACCACTAGCAGTAGC 546

QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 547 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTTCTAAGTGACATCAAG 606
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 607 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAAGTGAATTTGACTCAGGGCATCTGCCTGGAG 666
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 667 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAGACAGGGAGAGGCCCTGGCCGA 726
QY 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 727 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGATGCTGGGCCCATGATGCTCCCTGCTC 786
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 787 CTTGCCCAGTCTCAGGTGAGGCGCTCAGTGTCTACTGCTGCTTGGCCACAGAACAGAA 846
QY 241 IleSerSerLysLeuGlnLeuMetLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 847 ATTTCAGCAAACTCCAACTTATGAAAAGCACCAATCTGACCTGAAAAGCTGGGATC 906
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 907 CTAGATTTCACTCAGCAAGATGTTGCAAGCCACACAGAGCTATTCCCAAGACCCCTGATT 966
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 967 GCACCTGTCACCTCGGAGCCCTGCTGGCTGTCTTGGGCATCACTGGCTATTTCTCTGATG 1026
QY 301 Asn 301
Db 1027 AAT 1029
RESULT 2
LOCUS HUMCD34HS 2615 bp mRNA linear PRI 02-NOV-1993
DEFINITION Human CD34 mRNA, complete cds.
ACCESSION M81104.1
VERSION M81104.1 GI:180108
KEYWORDS CD34; hematopoietic stem cell surface antigen; sialomucin.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2615)
AUTHORS Simmons,D.L., Satterthwaite,A.B., Tenen,D.G. and Seed,B.
TITLE Molecular cloning of a cDNA encoding CD34, a sialomucin of human
hematopoietic stem cells
JOURNAL J. Immunol. 148, 267-271 (1992)
MEDLINE 92091783
FEATURES
source
1..2615
/organism="Homo sapiens"
/db_xref="taxon:9606"
294..1415
/codon_start=1
/product="CD34"
/protein_id="AAA03181.1"
/db_xref="GI:180109"
/translation="MPRGWTAICLLSLPSGFMSLDNGTATPELPTGCTFSNVSTNV
SVQETTPSTLIGSTSLHPVSOHQNEATNTITETTKFTSTSVIISVYGNNTSSVQSQT
ABIKCSGIREVKLTQICLEQNKTSSCAEFKDGRGLARVLGCEQADADAGAVCS
LLLAQSEVRPQCLLVLANRTEISSKLQMKKHQSDLKRLGLDFTEDDVASHQYSQ
KTLLALVTSALLVLTGYFLMNRSSWPTGERLGEDPYITENGGGQGYSSGPGTS
PEAQGKASVNRGAGKNGTQATSRNGHSARQHVAVDEL"
BASE COUNT 616 a 763 c 627 g 609 t
ORIGIN
Alignment Scores:

Pred. No.: 1,04e-91 Length: 2615
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-6 (1-301) x HUMCD34HS (1-2615)

```
QY 1 MetProArgGlyTrpThrAlaLeuSerLeuSerLeuProSerGlyPheMetSer 20
|||||
Db 294 ATCCCGGGGGCTGGACCGGCTTCTGCTGAGTTGCTGCTTCTGGGTTTCATGAGT 353
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 354 CTTGACAAACACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAATGTT 413
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 414 TCTACAAATGTATCTTACCAAGAACTACAAACCTAGTACCTTGGAGTACCGCTG 473
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
|||||
Db 474 CACCCCTGTCTCAACATGGCAATGAGGCCACAAACATCACAGAAACGACAGTCAA 533
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
Db 534 TTCACATCTACCTCTGTGATAACCTCAGTTTANGAAACACAAACCTCTCTGTCCAGTCA 593
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||||
Db 594 CAGACCTCTCTAATCAGCACAGTGTTCACCAACCCAGCCAAACGTTTCAACTCCAGAGACA 653
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
|||||
Db 654 ACCTTGAAAGCCTAGCTGTCCACCTGGAAATGTTTTCAGACCTTTCAACCCACTACGACTAGC 713
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 714 CTTGCAACATCTCCACTAAACCTATACATCATCTCTCTATCTCTAAGTACATCAAG 773
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
|||||
Db 774 GCAGAAATCAATGTTTCAGGCATCAGAGAGTGAATTTGACTCAGGGGCATCTGCCTGGAG 833
QY 181 GluAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
|||||
Db 834 CAAATTAAGACCTCCAGCTGTGGGAGTTTAAAGAGGACAGGGGAGAGGCGCTGGCCCGA 893
QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
|||||
Db 894 GTGCTGTCTGGGAGGAGCAGGTGATGCTGCTGGGCCAGGTATGCTCTCCTGCTC 953
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
|||||
Db 954 CTTGCCAGTCTGAGTGAGGCGCTCAGTGTCTACTGTGCTTGGCCAAACAGAACAGAA 1013
QY 241 IleSerSerLysLeuGlnLeuMetLysHisGlnSerAspLeuLysLysLeuGlyIle 260
|||||
Db 1014 ATTTCCAGCAACTCCAACTTATGAAGAGCACCATCTACCTGAAAGAGCTGGGGATC 1073
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
|||||
Db 1074 CTAGATTTCTACTGAGCAAGATGTTGCAAGCCACAGAGCTATTCCCCAAAAGACCTGATT 1133
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
|||||
Db 1134 GCAGTGCTACCTCGGGAGCCCTGCTGGCTGTCTTGGGCATCAGTGGCTATTTCCTGATG 1193
QY 301 Asn 301
|||
Db 1194 AAT 1196
```

RESULT 3
AX333753
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AX333753
Sequence 4262 from Patent WO0194629.
AX333753
AX333753.1 GI:18124472
human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)

AUTHORS
Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
gene sets

JOURNAL
Patent: WO 0194629-A 4262 13-DEC-2001;
Avalon Pharmaceuticals (US)

FEATURES
Location/Qualifiers
1..2657
/organism="Homo sapiens"
/db_xref="taxon:9606" 634 t

BASE COUNT 624 a 790 c 609 g 634 t

ORIGIN

Alignment Scores:
Pred. No.: 1,07e-91 Length: 2657
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-836-602-6 (1-301) x AX333753 (1-2657)

```
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 127 ATCCCGGGGGCTGGACCGGCTTCTGCTGAGTTGCTGCTTCTGGGTTTCATGAGT 186
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 187 CTTGACAAACACGGTACTGCTACCCAGAGTTTACCTACCCAGGGAACATTTTCAAATGTT 246
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 247 TCTACAAATGTATCTTACCAAGAACTACAAACCTAGTACCTTGGAGTACCGCTG 306
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
|||||
Db 307 CACCCCTGTCTCAACATGGCAATGAGGCCACAAACATCACAGAAACGACAGTCAA 366
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
Db 367 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACCTCTCTGTCCAGTCA 426
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||||
Db 427 CAGACCTCTGTATACGACAGTGTTCACCAACCCAGCCAGCTTTCAACTCCAGAGACA 486
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
|||||
Db 487 ACCTTGAAAGCCTAGCCTGTCCACCTGGAAATGTTTTCAGACCTTTCAACCCACTAGCCTAGC 546
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 547 CTTGCAACATCTCCCACTAAACCTATACATCATCTCTCTATCTCTAAGTACATCAAG 606
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
|||||
Db 607 GCAGAAATCAATGTTTCAGGCATCAGAGAGTGAATTTGACTCAGGGGCATCTGCCCTGGAG 666
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
|||||
```

Db	667	CAAAATAAGACCTCCAGCTGTCGGAGTTTAAGAAAGGACAGGGAGAGGGCTGGCCCGA	726
Qy	201	ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu	220
Db	727	GTGCTGTGTGGGAGAGCAGCTGATGCTGATGCTGGGCCAGGATGCTCCCTGCTC	786
Qy	221	LeuAlaGlnSerGluValargProGlnCysLeuLeuValLeuAlaAsnArgThrGlu	240
Db	787	CTTGCCAGCTCTGAGTGAGGCTCAGTCTCTACTGCTGTGCTTGGCCAAACAGAA	846
Qy	241	IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLeuGlyIle	260
Db	847	ATTTCCAGCAAACTCAACTTATGAAAGACCAATCTGACCTGAAAAGCTGGGGATC	906
Qy	261	LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeu	280
Db	907	CTAGATTTCAGTACAGAGATGTTGACGCCACAGAGCTATTTCCAAAGACCTGATT	966
Qy	281	AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeu	300
Db	967	GCACCTGGTCACCTCGGGAGCCCTGCTGCTGCTTTGGGCATCACTGGCTATTTCTG	1026
Qy	301	Asn 301	
Db	1027	AAT 1029	
RESULT 4			
S53911			
LOCUS	S53911	2657 bp mRNA linear PRI 08-MAY-1993	
DEFINITION	CD34-glycoprotein expressed in lymphohematopoietic progenitor cells (alternatively spliced, truncated form) [human, UT7, mRNA, 2657 nt].		
ACCESSION	S53911		
VERSION	S53911.1	GI:264768	
KEYWORDS	human UT7.		
SOURCE	human UT7.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 2657)		
TITLE	Nakamura,Y., Komano,H. and Nakauchi,H.		
JOURNAL	Two alternative forms of cDNA encoding CD34		
MEDLINE	Exp. Hematol. 21 (2), 236-242 (1993)		
REMARK	93146100		
FEATURES	GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 124009] from the original journal article. This sequence comes from Fig. 1AB.		
source	Location/Qualifiers		
gene	1..2657		
CDS	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	1..2657		
	/gene="CD34"		
	91..1077		
	/gene="CD34"		
	/note="glycoprotein expressed in lymphohematopoietic progenitor cells; This sequence comes from Fig. 1AB"		
	/codon_start=1		
	/product="CD34"		
	/protein_id="BAB25223.1"		
	/db_xref="GI:264769"		
	/translation="MLVRRGARAGPRMGWTLCLLSLLPSGFMSLDNRGTATPELP		
	TQGFNSVQSVQSTVSVFTPTSLGSLHPVSQHGNEATTNITETVTKFTSTVTSV		
	YGNTPSVQSVQSTVSVFTPTSLGSLHPVSQHGNEATTNITETVTKFTSTVTSV		
	YTNSSPILSDIKIKGSGIREVKLTGICILEONKTSCEAEFKDRGEGGLARVLGEE		
	QADADAGVCSLLLAQSEVRPOCLLLVLANRREISSKLQIMKKHQSDLKKLGILDT		
	EQDVASHQSTVSKTLLIALVTSGALLAVGITGTFMLNRRSWSPTGERLELEP"		
BASE COUNT	624 a 790 c 609 g 634 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1..07e-91	Length:	2657
Score:	1509.00	Matches:	301

Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0
US-09-836-602-6 (1-301) x S53911 (1-2657)			
Qy	1	MetProAlaGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer	20
Db	127	ATGCCCGGGGCTGGACCGCGCTTGGCTTGTGAGTTTGGCTTCTGGGTTCTATGAGT	186
Qy	21	LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
Db	187	CTTGACAAACAACGGTACTACCCACAGATTACCTACCCAGGGAACATTTTCAATGTT	246
Qy	41	SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu	60
Db	247	TCTACAAATGTATCTTACCAAGAACTACAACACCTAGTACCTTTGGAAGTACCA	306
Qy	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys	80
Db	307	CACCTGTGTCTACATGGCAATGAGGCCACAAACATCACAGAACCACAGTCAAA	366
Qy	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer	100
Db	367	TTCACATCTACTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA	426
Qy	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	427	CAGACCTCTGTAATCAGACACAGTGTTCACCAACCCAGCAACGTTTCAACTCCAGAGACA	486
Qy	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrSerThrSer	140
Db	487	ACCTTGAAGCCCTAGCCTGTACCTGGAATGTTTCAGACCTTTCAACCACTAGCACTAGC	546
Qy	141	LeuAlaThrSerProThrLysProThrThrSerSerSerProIleLeuSerAspIleLys	160
Db	547	CTTGCAACATCTCCACATAAACCCCTATACATCATCTTCTCTATCTTAACTAGTACATCAAG	606
Qy	161	AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu	180
Db	607	GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCAGGGCATCTGCCTGGAG	666
Qy	181	GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg	200
Db	667	CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAGGACAGGGAGAGGGCTGGCCCGA	726
Qy	201	ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu	220
Db	727	GTGCTGTGTGGGAGGAGCAGGCTGATGCTGATGCTGGGGCCAGGATGCTCCTGCTC	786
Qy	221	LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu	240
Db	787	CTTGCCAGCTCTGAGGTGAGGCCTCAGTGTCTACTGTGTGGCTTGGCCCAACAGAACAGAA	846
Qy	241	IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLeuGlyIle	260
Db	847	ATTTCCAGCAAACTCAACTTATGAAAAGACCAATCTGACCTGAAAAGCTGGGGATC	906
Qy	261	LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeu	280
Db	907	CTAGATTTCAGTACGCAAGATGTTTGAAGCCACAGAGCTATTTCCAAAGACCTGATT	966
Qy	281	AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet	300
Db	967	GCACCTGGTCACCTCGGGAGCCCTGCTGCTGCTTTGGGCATCACTGGCTATTTCTGATG	1026
Qy	301	Asn 301	
Db	1027	AAT 1029	
RESULT 5			
CFU49457			

LOCUS CFU49457 2956 bp mRNA linear MAM 21-JAN-1997
DEFINITION Canis familiaris hematopoietic progenitor cell marker CD34 mRNA,
 complete cds.
ACCESSION U49457
VERSION U49457.1 GI:1224105
KEYWORDS
SOURCE dog.
ORGANISM Canis familiaris
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE
AUTHORS McSweeney,P.A., Rouleau,K.A., Storb,R., Bolles,L., Wallace,P.M.,
 Beauchamp,M., Krizanac-Bengez,L., Moore,P., Sale,G., Sandmaier,B.,
 de Revel,T., Appelbaum,F.R. and Nash,R.A.
TITLE Canine CD34: cloning of the cDNA and evaluation of an antiserum to
 recombinant protein
JOURNAL Blood 88 (6), 1992-2003 (1996)
MEDLINE 96420219
REFERENCE 2 (bases 1 to 2956)
AUTHORS McSweeney,P.A.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-1996) P.A. McSweeney, Transplant Biology, FHCR,
 1124 Columbia St, Seattle, WA 98104, USA
FEATURES
 source
 1..2956
 Location/Qualifiers
 /organism="Canis familiaris"
 /db_xref="taxon:9615"
 /tissue_type="ML2 myelomonocytic leukemia"
 1..2956
 /gene="CD34"
 260..1429
 /gene="CD34"
 /note="type 1 transmembrane protein"
 /codon_start=1
 /product="hematopoietic progenitor cell marker CD34"
 /protein_id="AAB41055.1"
 /db_xref="GI:1224106"
 /translation="MLAGRGAGGLPRGWTALCLLSLLPGFTNTETVTPPTVT
 STEIMASVENTSKREALTLPSTGTTLYSVSDSSGTAISERTVHVSTSEITLT
 PGTMNSVQSOTSLAITVFTPTNFSTSVILEPSLLPGNSDPPYNSLTVSTPEY
 YTSLSPTSRNDPTSTIGKICSGVKEVKNQICLEINETSCEDEPKKNEEKLTQ
 VLCEPAEAGAGVCSLLAQSEVRPHCLLVLANKEFLSKLQRKHQSDLKLI
 RDTFQDQSHQSVSRKTLIALTVSGILLVLGTTGYFLMNRSSWPTGERLGEDPY
 TENGGGQGVSSGPGVPEAQGKASVNRGPQNGTQATSRNGHSARQHWADTEL"
 BASE COUNT 625 a 864 c 815 g 652 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5.67e-51 Length: 2956
 Score: 899.00 Matches: 196
 Percent Similarity: 72.31% Conservative: 26
 Best Local Similarity: 63.84% Mismatches: 77
 Query Match: 59.58% Indels: 8
 DB: 4 Gaps: 2
 US-09-836-602-6 (1-301) x CFU49457 (1-2956)
 QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
 Db 296 CTGCCCGGGGCTGGACCGCGCTCTGCTGCTCAGTCTGCTGCCCTTGTGGTTCACAAAC 355
 QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 Db 356 ACAGAACCGTGATTACTCCTCCACACAGTGCCCAACCTCCACAGAAATAATGTCAGCTGT 415
 QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
 Db 416 TCTGAGAATACATCCAAACGGGAGCCATCACACTAACTCTCTCTGGAAGTACACCTG 475
 QY 61 HisProValSerGlnHisGlyAsnGluAlaThrAsnIleThrThrValLys 80
 Db 476 TACTCTGTCTCTCAAGACAGCAGTGGGACACAGCAACCATCTCAGAGACTACAGTCCAT 535

QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 Db 536 GTACACATCTACCTCTCAGATCACCTAAGCCCTGGGACCATGAACCTCTCTCTCAGTCG 595
 QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 Db 596 CAGACCTCTTTAGCTATCACGGTATCTTTTACCCCAACCAACTTTTCAACTTCAAGTGTG 655
 QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
 Db 656 ACCTTGGAGCCCGAGCTGCTACTTGGAAATGTTGGATCCCTCCCTTACACAGCACACGAG 715
 QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerPro----- 154
 Db 716 CTGTGCACATCCCCACCGAATATTATACATCATTCTCTCTACCCCAAGTAGAATGAC 775
 QY 155 IleLeuSerAspIleLysAlaGluIleLysCysSerGlyIleAargGluValLysLeuThr 174
 Db 776 ACCCCAAAGTACCATCAAGGAGAAATCAAAATGTTCCGGAGTCAAGAGAGTGAATTTGAAC 835
 QY 175 GlnGlyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArg 194
 Db 836 CAAGGTATCTGCTAGAGCTAAATGAGACTCCAGCTGTGAGGACTTTAAGAAAGATAAC 895
 QY 195 GlyGluGlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAla 214
 Db 896 GAAGAAAACTGACCAAGTCTCTGT-----GAGAAGGAGCCAGCTGAGGCTGGGGCC 949
 QY 215 GlnValCysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuVal 234
 Db 950 GGGGTGTGCTCCTGCTCTGCGCCAGTCTGAGGTGAGGCTCAGCTGCTGCTGCTGCTG 1009
 QY 235 LeuAlaAsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAsp 254
 Db 1010 TTGGCCCAACAAACAGACACTTTTCAGTAACTCCAACTTCTGAGAAAGCACCAGTCTGAC 1069
 QY 255 LeuLysLysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyr 274
 Db 1070 CTGAAAAGCTGGGATCCGAGACTTCACTGAACAGATGTTGGGAGCCACACAGACTAT 1129
 QY 275 SerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIle 294
 Db 1130 TCCCGCAAGACCTGATTGCACCTGCTGCTCAGGATCTGCTGCTGCTGCTGCTGCTG 1189
 QY 295 ThrGlyTyrPheLeuMetAsn 301
 Db 1190 ACTGTTACTTCTCTGATGAC 1210
 RESULT 6
 AF461503 2690 bp mRNA linear MAM 24-JAN-2002
 LOCUS AF461503
 DEFINITION Sus scrofa CD34 antigen mRNA, complete cds.
 ACCESSION AF461503
 VERSION AF461503.1 GI:18308137
 KEYWORDS
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
AUTHORS Sun,J. and Butler,J.E.
TITLE Cloning, sequencing and expression of swine CD34 gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2690)
AUTHORS Sun,J., Wang,J. and Butler,J.E.
TITLE Direct Submission
JOURNAL Submitted (20-DEC-2001) Microbiology, University of Iowa, 51 Newton
 Road, Iowa City, IA 52242, USA
FEATURES
 source
 1..2690
 Location/Qualifiers
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /tissue_type="mesenteric lymph nodes"


```

Db 676 TGCTCCTGCTTCTAGCCCAAGTCTGAGGTTAGCCCTGAGTGTTGCTGATGGTCTGGCC 735
QY 237 AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
||| |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db 736 AATAGCACAGAACTCCACGAACTCCAGCTTATGGAAAGACCAATCTGACTTGAGA 795
QY 257 LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGln 276
||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||::: ||| |||||:::
Db 796 AAGCTGGGATCCAAAGCTTCAATAACAAAGATATCGGAGCCACCAAGAGCTATTCCCGA 855
QY 277 LysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGly 296
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 856 AAGACTCTTATTGATGGTCACTCTGGAGTCTGCTGGCCACTTGTGGGCACCACTGGT 915
QY 297 TyrPheLeuMetAsn 301
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 916 TATTTCCTGATGAAC 930

RESULT 9
S69299
LOCUS
DEFINITION S69299 1149 bp mRNA linear ROD 03-JUN-1994
[mice, stromal cell line PA-6, mRNA partial, 1149 nt].
ACCESSION S69299
VERSION S69299.1 GI:495713
KEYWORDS Mus sp. stromal cell line PA-6.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1149)
AUTHORS Suda,J., Sudo,T., Ito,M., Ohno,N., Yamaguchi,Y. and Suda,T.
TITLE Two types of murine CD34 mRNA generated by alternative splicing
JOURNAL Blood 79 (9), 2288-2295 (1992)
MEDLINE 92239883
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gbbseq 145456] from the original journal article.
This sequence comes from Fig. 3.
FEATURES
source
1. .1149
/organism="Mus sp."
/db_xref="taxon:10095"
1. .1149
/partial
/gene="CD34"
1. .1149
/gene="CD34"
/note="cell surface antigen; This sequence comes from Fig.
3"
/codon_start=1
/product="CD34"
/protein_id="AAB22108.1"
/db_xref="GI:495714"
/translation="MQVHRDTRAGLLLPWRWVALCMLSLHLLNLTSAATSTSGIS
TPSPNVEENITSSIGSTSHYLIYQDSKSTPAISETMVTNFTVGTGIPSGSTPH
TFSPQSTPLPTSDISISTSEMTKSLPSINVDYSPNNSPFMTSPTEPYAT
SSSAATKGLTCKGIRVRLAOGICLSESPACEFKKEDLDLILCEKEAE
ADAGASVCSLLLAQSEVRPECLMLVLANSTELPSKLOLMEKHQSDRLKIGIOENKQD
IGSHOSYRKTLIALVTSGVLLAIGTGYFLNRRNSPTGERLGEDPYPTENGGO
GYSSGPGASPEPQGANVTRGAQENGSTGQATSRNGHSARQHVVADEL"
BASE COUNT 307 a 293 c 285 g 264 t
ORIGIN

Alignment Scores:
Pred. No.: 2,9e-45 Length: 1149
Score: 805.50 Matches: 177
Percent Similarity: 71.15% Conservative: 40
Best Local Similarity: 58.03% Mismatches: 77
Query Match: 53.38% Indels: 11
DB: 10 Gaps: 4

US-09-836-602-6 (1-301) x S69299 (1-1149)

```

```

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
::: ||| ||| ||||| |||||::: |||||
Db 37 CTGCCATGGCGTGGGTAGCTCTCTGCCATGATGAGTCTGCTG-----CAT 81
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
||||| |||||::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 82 CTAATAAATCTTGACTCTCTACCCAGGAGACTTCTACAAAGGAATATCCCATCAGTT 141
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 142 CCTACCAATGAGTCTCTGTGAGGAAATATACATCTAGCATCCTCGAAGTACCAAC 201
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
::: ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 202 TACTTGATCTATCAGGACAGCAGTAAGACACCAACCAAGCATCTCAGAGACTATGGTCAAC 261
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262 TTTACAGTTACCTCTGGGATCCCTTCAGGCTCTGGAACTCCACACACTTTTTCACAACCA 321
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 322 CAGACTTCCCAACTGGCATCTGCTACTACTACTCTCAGACAGATATTTCCACTTCAGAGATG 381
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 382 ACCTGGAAGTCCAGCTGCCATCTATAAATGTTTCTGATTATTCGCTATATAATAGCAGC 441
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 141 Leu--AlaThrSerProThrLysProTyr-----ThrSerSerSerProIleLeu 156
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 442 TTTGAGATGACATCACCCACCGAGCATATGCTTACACATCATCTTCTGCTCG----- 495
QY 157 SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 496 AGTGCCATTAAGGAGAAATCAATGCTCTGGAATCCGAGAAGTGAGGTGGCCAGG 555
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 177 IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 556 ATCTGCTGGAACTAAAGTGAAGCATCTAGTTGTGAGGAGTTTAAAGAAAGAAAGGAGAA 615
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 197 GlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnVal 216
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 616 GATCTAATTCAAATACTGTGTGAAAGGAGGAGGCTGAGCTGATGCTGTGCTAGTGTC 675
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 217 CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 676 TGCTCCTGCTTCTAGCCAGTCTGAGGTAGGCTGAGTGTGTTGCTGATGGCTTTGGCC 735
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 237 AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 736 AATAGCACAGAACTTCCAGCAAACTCCAGCTTATGGAAGACCAATCTGACTTGAGA 795
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 257 LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGln 276
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 796 AAGCTGGGATCCAAAGCTTCAATAACAGATATCGGAGCCACCAAGACTATTCCCGA 855
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 277 LysThrLeuIleAlaLeuValThrSerGlyAlaLeuAlaValLeuGlyIleThrGly 296
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 856 AAGACTCTTATTGCAATGGTCACTCTGAGGTTCTGCTGGCCATCTTGGCCACCACTGGT 915
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 297 TyrPheLeuMetAsn 301
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 916 TATTTCCTGATGAAC 930

RESULT 10
S69293
LOCUS
DEFINITION S69293 1260 bp mRNA linear ROD 03-JUN-1994
[CD34-stem cell antigen (mice, bone marrow, mRNA, 1260 nt)].
ACCESSION S69293
VERSION S69293.1 GI:495708
KEYWORDS Mus sp. bone marrow.
SOURCE Mus sp.
ORGANISM Mus sp.

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1260)
Brown, J., Greaves, M.F. and Molgaard, H.V.

The gene encoding the stem cell antigen, CD34, is conserved in mouse and expressed in haemopoietic progenitor cell lines, brain, and embryonic fibroblasts
Int. Immunol. 3 (2), 175-184 (1991)

91223042

GenBank staff at the National Library of Medicine created this

entry [NCBI gibbsq 145442] from the original journal article.

This sequence comes from Fig. 2.

Location/Qualifiers

1..1260

/organism="Mus sp."

/db_xref="taxon:10095"

1..1260

/gene="CD34"

24..1172

/gene="CD34"

/note="stem cell antigen; This sequence comes from Fig. 2"

/codon_start=1

/product="CD34"

/protein_id="AAB19246.1"

/db_xref="GI:495709"

/translation="MOVHRTDRLGLLPWRVALCLMLLNLLNLTTSATTTSTQGIS

PSVPTNESVRENTTSSVSHVLYDSSKTPALSETVNFVTWSPGIPSGSTPH

TFQSPQSPGILPTTSDSISTSEMTWKSLSIPNSDYSFNSSFWMTPTPEYAT

SSASAKIEKICSGIREVRLAAGICLSEASEEKEFKEDGDIQLICKEEAE

ADAGASVCLLQAQSEYRPECLMLANSTELPSKLQMEKHQSDRLKLGTSFNKQD

IGSHQSRKLTALGVLLAILGTGYFLMNRNWSPTGERLGDPTVTENGSGQ

GYSSGSGPETQCKANVTGRGAQENGTQATSRNGHSARQHVVDTEL"

BASE COUNT 325 a 321 c 320 g 294 t

ORIGIN

Alignment Scores:

Pred. No.: 3.26e-45 Length: 1260
Score: 805.50 Matches: 177
Percent Similarity: 71.15% Conservative: 40
Best Local Similarity: 58.03% Mismatches: 77
Query Match: 53.38% Indels: 11
DB: 10 Gaps: 4

US-09-836-602-6 (1-301) x S69293 (1-1260)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 60 CTGCCATGGCGCTGGTGGTACCTCTGCCTGATGAGTCTGCTG-----CAT 104
Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 105 CTAATAACTTCACTCTGCTACCGAGAGACTTCTACACAGGAATATCCCATCAGTT 164
Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 165 CCTACCAATGAGTCTGTTGAGGAAATATCATCATGACCTCCCTGGAAGTACCAGCCAC 224
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 225 TACTTGATCTATCAGACAGCAGTGAAGACACACACGAGCATCTCAGAGATGTTGTCAC 284
Qy 81 PheThrSerThrValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 285 TTTACAGTTACTCTGGGATCCCTTCAGGCTCTGGAATCCACACATTTTTCACACCA 344
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 345 CAGACTTCCCAACGCGCATAGTCTACTTCTCAGACAGTATTTCCACITTCAGAGATG 404
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 405 ACCTGGAAGTCCAGCTGCCATCTATAATGTTCTGATTTATTCGCTTATAATAATAGCAGC 464

Qy 141 Leu---AlaThrSerProThrLysProTyr-----ThrSerSerSerProfileLeu 156
Db 465 TTTGAGATGACATCACCACCGAGCCATATGCTTACACATCATCTTCTGCTCCG----- 518
Qy 157 SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
Db 519 AGTGCCATTAGGAGGAATCAATGCTCTGGAATCCGAGAAGTGAGGTTGGCCAGGTT 578
Qy 177 IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
Db 579 ATCTGCCCTGGAACTAAGTGAAGCATCTAGTTGTGAGGAGTTTAAAGAGGAAAGGAGAA 638
Qy 197 GlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyValaGlnVal 216
Db 639 GATCTAATTCAAATACTGTGTCAAAGAGGAGGCTGAGGCTGATGCTGGTGTAGTGTG 698
Qy 217 CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
Db 699 TGCTCCCTGCTTCTAGCCACGCTCAGGTTAGGCTGAGTGTGCTGCTGCTGCTGCTG 758
Qy 237 AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
Db 759 AATAGCACAGAACTTCCAGCAAACTCCAGCTTATGGAAGACCACTGACTGACTGAGA 818
Qy 257 LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGln 276
Db 819 AAGCTGGGATCCAAAGCTTCAATAAACAAGATATCGGAGGCCACAGACTATTTCCGA 878
Qy 277 LysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGly 296
Db 879 AAGACTCTTATTGCACTTGGTCACTCTGAGTCTGCTGGCCACTCTTGGCCACCACTGT 938
Qy 297 TyrPheLeuMetAsn 301
Db 939 TATTTCTGATGAAC 953
RESULT 11
BC006607 2398 bp mRNA linear ROD 12-JUL-2001
LOCUS Mus musculus, Similar to CD34 antigen, clone MGC:11720
DEFINITION IMAGE:3966337, mRNA, complete cds.
ACCESSION BC006607
VERSION BC006607.1 GI:13879273
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2398)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NTH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalon@bcm.tmc.edu
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 17 Row: n Column: 8

This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES		Location/Qualifiers	
source		1..2398	
		/organism="Mus musculus"	
		/db_xref="taxon:10090"	
		/clone="MGC:11720 IMAGE:3966337"	
		/tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."	
		/clone_lib="NCLCGAP_Mam1"	
		/lab_host="DH10B"	
		/note="Vector: pCMV-SPORT6"	
CDS		62..1210	
		/codon_start=1	
		/product="Similar to CD34 antigen"	
		/protein_id="AAH06607.1"	
		/db_xref="GI:13879274"	
		/translation="MOVHRDTRAGLLPWRVALCLMSLLHNLNLTSAFTTSTOGIS PSVPTNESVEENITSSIGPSTSHYLIYQDSKTPAISRTMVFNTVTSIPSGSGTPH TFSQPTSGTGLPTTSDISISTSEMTKSSLPINVSIDYSPNNSFEMTSPTEPVAYT SSASPAKGEIKGSCIREVRLAOGICILESPASCEEFKEKGDILQILCEKEAE ADAGASVSLLAQSEVRECLLMVLANSTELPSKLQIMEKHQSDLRKLIGISFNKQD IGHSQYSKRTIALIVTSGVLLAILGTGFLMNRKMSPTGERLGEDPYITENGCGQ GYSSGPGASPEQTQKANVTRGAENGTCQATSRNGHSARQHVVAADTEL"	
BASE COUNT		623 a 599 c 551 g 625 t	
ORIGIN			
Alignment Scores:			
Pred. No.:	7.42e-45	Length:	2398
Score:	805.50	Matches:	177
Percent Similarity:	71.15%	Conservative:	40
Best Local Similarity:	58.03%	Mismatches:	77
Query Match:	53.38%	Indels:	11
DB:	10	Gaps:	4
US-09-836-602-6 (1-301) x BC006607 (1-2398)			
QY	1	MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer	20
Db	98	CTGCCATGGCGCTGGGTAGCTCTCGCTGATGAGTCTGCTG-----CAT	142
QY	21	LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
Db	143	CTAATAACTGTACTCTGCTACCCAGGAGACTTCTACACAAGGAATATCCCATCAGTT	202
QY	41	SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu	60
Db	203	CCTACCAATGAGTCTGTTGAGGAAAAATATCACAATCTAGCATCCTCGGAAGTACGACC	262
QY	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys	80
Db	263	TACTTGATCATCAGGACAGCAGTAAAGACCACACGACCATCTCAGAGACTATGGTCAAC	322
QY	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer	100
Db	323	TTTACAGTTACCTCTGGGATCCCTTCAGGCTCTGGAATCCACACACTTTTTCACAACCA	382
QY	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	383	CAGACTTCCCAACTGGCACTACTGCTACTTTCAGACAGTAGTATTTCCACTCAGAGATG	442
QY	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer	140
Db	443	ACCTGGAAGTCCAGCCTGCCATCTATAAATGTTTCTGATTATTCGCTTAATAATAGCAGC	502
QY	141	Leu---AlaThrSerProThrLysProTyr-----ThrSerSerSerProIleLeu	156
Db	503	TTTGAGATGACATCACCCAGGACCATATGCTTACACATCATCTTCTGCTCG-----	556
QY	157	SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly	176
Db	557	AGTGCCATTAAAGGAGAAATCAATGCTCTGGAATCCGAGAGTAGTGAGGTGGCCCGGGT	616

QY	177	IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu	196
Db	617	ATCTGCTGGAACTAAGTGAAGCATCTAGTTGTGAGGAGTTTAAAGAGGAAAGGAGAA	676
QY	197	GlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnVal	216
Db	677	GATCTAATTCAAATCTGTGTGAAAGAGGAGGCTGAGGCTGATGCTGTGTAGTGTCT	736
QY	217	CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla	236
Db	737	TGCTCCTCTGCTTCTAGCCAGTCTGAGTTAGGCTGAGTGTGCTGATGCTGTGGCC	796
QY	237	AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys	256
Db	797	AATGACAGACAACCTCCACAGCAATCCAGCTTATGGAAGACCAATCTGACTTGAGA	856
QY	257	LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGln	276
Db	857	AGCTGGGATCCAAAGCTTCAATAACAAGATATCGGAGCCACAGAGCTATCCCGA	916
QY	277	LysThrIleuIleAlaLeuValThrSerGlyAlaLeuAlaValLeuGlyIleThrGly	296
Db	917	AGACTCTTATTGCTTGGTGCACCTCTGGAGTCTGCTGCCATCTTGGCCACCACTGGT	976
QY	297	TyrPheLeuMetAsn	301
Db	977	TATTCTGATGAAC	991
RESULT 12			
AF202879			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
gene			
CDS			
BASE COUNT			
ORIGIN			

```
Alignment Scores: 1.78e-40 Length: 581
Pred. No.: 728-00 Matches: 147
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 48.24% Gaps: 0
DB: 9

US-09-836-602-6 (1-301) x AF202879 (1-581)
QY 155 IleLeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThr 174
DB 2 ATCTAAGTGACATCAAGCAGAAATCAATGTTTCAGGCATCAGAGAAGTGAATGACT 61
QY 175 GlnGlyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArg 194
DB 62 CAGGCAATCTGCCTGGAGCAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAGGACAGG 121
QY 195 GlyGluGlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAla 214
DB 122 GGAGAGGCTTGGCCAGTGTCTGTGGGAGGAGCAGGCTGATGCTGATGCTGGGGCC 181
QY 215 GlnValCysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuVal 234
DB 182 CAGGTATGCTCCCTGCTCTTGGCCAGCTGAGGTGAGGCTCAGTCTACTGCTGTC 241
QY 235 LeuAlaAsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAsp 254
DB 242 TTGGCCAAACAGAGAGAAATTTCCAGCAAACTCCAACTTATGAAAAAGCACCATCTGAC 301
QY 255 LeuLysLysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyr 274
DB 302 CTGAAAAGCTGGGATCTCTAGATTTCTACGTGAGCAAGATGTTGCAAGCCACAGAGCTAT 361
QY 275 SerGlnLysThrLeuIleAlaLeuValThrserGlyAlaLeuLeuAlaValLeuGlyIle 294
DB 362 TCCCAAAAGACCCGATTTGCACTGGTCACCTCGGAGCCCTGCTGCTGCTGTTGGGCATC 421
QY 295 ThrGlyTyrPheLeuMetAsn 301
DB 422 ACTGGCTATTTCCTGATGAAT 442

RESULT 13
AB021662 Bos taurus mRNA 2453 bp mRNA linear MAM 28-NOV-2001
LOCUS
DEFINITION Bos taurus mRNA for CD34, complete cds.
ACCESSION AB021662
VERSION AB021662.1 GI:4996448
KEYWORDS CD34.
SOURCE Bos taurus (strain:Holstein) CDNA to mRNA.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
1 (sites)
Zhou,J.H., Hikono,H., Ohta,M. and Sakurai,M.
Cloning of bovine CD34 cDNA
J. Vet. Med. Sci. 63 (9), 1051-1053 (2001)
21519980
REFERENCE
2 (bases 1 to 2453)
Sakurai,M., Zhou,J., Hikono,H. and Ohta,M.
Direct Submission
Submitted (21-Dec-1998) Michiharu Sakurai, National Institute of
Animal Health, Laboratory of Immunobiology; 3-1-1 Kannondai,
Tukuba, Ibaraki 305-0856, Japan (E-mail:yukari@niah.affrc.go.jp,
Tel:81-298-7790, Fax:81-298-7880)
FEATURES
source
1..2453
/organism="Bos taurus"
/strain="Holstein"
/db_xref="taxon:9913"
50...1198
gene
```

```
CDS
/gene="CD34"
50...1198
/note="cluster of differentiation antigen 34"
/codon_start=1
/product="CD34"
/protein_id="BAA78476.1"
/db_xref="GI:4996449"
/translation="MLGRGARAGRCWVGRWATLCLLLPSMEFTQTQPMATATVTS
ASPEISSPVPTFTTDRWSMNSSTFRSTSLYDVSQDSNGTTAVTLVPASTDSTSTSET
LPASGAVNSVQQTSLATVTSSTLFSATSEATLQPTFPNGNISLNSSTPVTSS
ITSPSPSTQSLKSEIKCSRVKEVLAHSICLKNLTSSCEFKNNKEIKTLGFLC
OKKQAEVCSLLIAQSEVRPQCLLLVLTNRTESSKIKLLEHWSDLREMGIEDISBEA
VSSHQSYSRKTLIALVTSGLLAILTTCYFLMNRBSWSPTRGERLGEDPYVVENGGQ
GYSSGPEASPEAOGKALVNRNGSTGQATSRNGHSRQRVVVADTEL"
BASE COUNT 522 a 708 c 654 g 569 t
ORIGIN
Alignment Scores: 2.61e-38 Length: 2453
Pred. No.: 707-50 Matches: 167
Score: 64.95% Conservative: 35
Percent Similarity: 53.70% Mismatches: 86
Best Local Similarity: 46.89% Indels: 23
Query Match: 4 Gaps: 5
DB: 4
US-09-836-602-6 (1-301) x AB021662 (1-2453)
QY 1 MetProArgGlyThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
DB 86 ATGCTGGGGCTGACCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 133
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
DB 134 -----ATGGAATTCACACAGACACCCATGCGAGCTACTGTCACAGTGTCA 184
QY 41 SerThrAsnValSerTyrGlnGluThrThrThr-----Pro 52
DB 185 AGCCCAAAATATCTTCACCTGTCCCTACAACTACATCGACTGGAGNAGCATCACTCA 244
QY 53 SerThrLeuGlySerThrSerLeuHisProValSerGln-----HisGlyAsnGluAla 70
DB 245 AGTACTTTTAGAAGCACCAGCCTCTACGATGCTCTCAGGACAGCAATGGGACCACAGA 304
QY 71 ThrThrAsnIleThrGluThrThrValLysPheThrSerThrSerValIleThrSerVal 90
DB 305 GTCACCTTAGTTCACGCGCTACAGACAGTTTCACGTTCTACCTCTGAGACCCCTCCAGCC 364
QY 91 TyrGlyAsnThrAsnSerSerValGlnSerGlnThrSerValIleSerThrValPheThr 110
DB 365 TCTGGGGCCGTGAACCTCTCTGTCCTCAGACAGACCTCTTTAGCCACCACCGTCTCTCT 424
QY 111 ThrProAlaAsnValSerThrProGluThrThrLeuLysProSerLeuSerProGlyAsn 130
DB 425 ACCCTCATCAGCTTTGGCAGCTTCAGAGGCGACTCTGCGAGCCAGCACTTTTCTGGAAT 484
QY 131 ValSerAspLeuSerThrThrSerThrSerLeuAlaThrSerProThrLysProThrThr 150
DB 485 ATTTTCGGATTCCCTGTCAATATAGTACCAGCCCTGTGACTTCTCCATTAATTAACCTCTCCA 544
QY 151 SerSerSerProIleLeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGlu 170
DB 545 TCATTTTCTCTACCCAAAGTATCTCAAGAGTGAATAAATGTTTCCAGAGTCAAGAA 604
QY 171 ValLysLeuThrGlnGlyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPhe 190
DB 605 GTGAAATTGGCCCAACAGCATCTGCTGAAAGCTAAATGAGACCTCCAGCTGTGAGAGTTT 664
QY 191 LysLysAspArgGlyGluGlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAla 210
DB 665 AAGAAGAACATGAAGAAACTGACCGGATTCCTGTGTGTCAGAGAAACAG-----715
QY 211 AspAlaGlyAlaGlnValCysSerLeuLeuAlaGlnSerGluValArgProGlnCys 230
```

```

Db 716 -----GCCAGGTGCTCTCTGCTCCGCGCCAGTCTGAGGTGAGGCTCAGTGC 766
|||||
Qy 231 LeuLeuValLeuAlaAsnArGThrGluIleSerSerLysLeuGlnLeuMetLysLys 250
|||||
Db 767 CTGCTGCTGCTCTGACCAACAGGACAGAAAGTTCACGCAAGATCAAACTCTCGGAAGAG 826
|||||
Qy 251 HisGlnSerAspLeuLysLysLeuGlyIleLeuAspPheThrGluGlnAspValAlaLaser 270
|||||
Db 827 CACTGCTGCTGACCTGAGAGAGAGTGGGCATCGAGGACATCTCTGAAGAAGCTGTCACGAGC 886
|||||
Qy 271 HisGlnSerTyrosSerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuAla 290
|||||
Db 887 CACCAGAGCTATTCGCGAAGACCTTGATTCGACTGCTCACCTCGGGGATCTCTGCTGGCC 946
|||||
Qy 291 ValLeuGlyIleThrGlyTyrosPheLeuMetAsn 301
|||||
Db 947 ATCTTGATCACCACGCTGCTATTTCTTGATGAAC 979
|||||
RESULT 14
AF259378 2420 bp mRNA linear MAM 14-JUN-2000
LOCUS
DEFINITION
ACCESSION AF259378
VERSION AF259378.1 GI:8515725
KEYWORDS
SOURCE
ORGANISM
Capra hircus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactylia; Ruminantia; Pecora; Bovidae;
Bovidae; Caprinae; Capra.
REFERENCE
AUTHORS Leipprandt,J.R., Anson,D.S. and Jones,M.2.
TITLE Cloning and sequence analysis of caprine CD34 cDNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2420)
AUTHORS Leipprandt,J.R., Anson,D.S. and Jones,M.2.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2000) Biochemistry, Michigan State University,
Room 519 Biochemistry, East Lansing, MI 48824, USA
FEATURES
source
1..2420
/oranism="Capra hircus"
/db_xref="taxon:9925"
<1..1069
/note="cell-surface antigen"
/codon_start=2
/product="CD34"
/protein_id="AA76145.1"
/db_xref="GI:8515726"
/translation="PSMEFTTAVTGNATVSSNPGISSVPVPTTTSYKMSNMSSTFRST
SLYNVQSDNGTAVTSVPASTVNFSTASSETLPASGAVNSVQPTSLATTSSTLIS
FATSEATLQPTFPNTSDSYNSTSPVTSINTSPSPQNTILKTEIKSRVKEV
LAHSTCLKNTSNCEEPKNNEEKLRYLCQEQAEVCSLLAQSEVKPQCLLLVLT
NRTSESSKRLRLQSDLRMGIDISEEDVSSHQSYSRKTLALVYSGILLAILIT
TCYFLMNRSSWSPAGERGDPYVVGNGGGQYSGSEASPEAQGKATVNRGAQENGT
GQATSRNGHSARQRVADTEL"
BASE COUNT 565 a 700 c 590 g 565 t
ORIGIN

Alignment Scores:
Pred. No.: 1.39e-36 Length: 2420
Score: 681.50 Matches: 152
Percent Similarity: 67.74% Conservative: 37
Best Local Similarity: 54.48% Mismatches: 81
Query Match: 45.16% Indels: 9
DB: 4 Gaps: 3

US-09-836-602-6 (1-301) x AF259378 (1-2420)

Qy 25 GlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnVal 44
|||||
Db 35 GGTATATGCCACAGTGTCA---TCCAAACCCAGGAATATCTTACCTGCTCCCTCAACACTAGC 91
|||||

```

```

Qy 45 SerTyrosGlnGluThrThrProSerThrLeuGlySerThrSerLeuHisProValSer 64
|||||
Db 92 TCTACTGGAAGAAGCATGAACCACTACTTTTAGAAGCACCACGCTCTACATGTCTCT 151
|||||
Qy 65 Gln-----HisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLysPheThr 82
|||||
Db 152 CAGGACAGCAATGGGACACAGCAGTCAGTCTAGTCCAGCGTCTACAGTCAATTTTCACG 211
|||||
Qy 83 SerThrSerValIleThrSerValTyrosGlyAsnThrAsnSerSerValGlnSerGlnThr 102
|||||
Db 212 TCTGCTCTGAGACCCCTCCAGCCTCTGGAGCGCGTGAACCTCTTCTGTCCAACACAGACC 271
|||||
Qy 103 SerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThrThrLeu 122
|||||
Db 272 TCTTTAGCCACCACGCGCATCTTCTACCTCATCAGCTTTGGCAGCTTCAGAGGCGACTCTG 331
|||||
Qy 123 LysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSerLeuAla 142
|||||
Db 332 CAGCCACGACGCTTCTCGAAATATTTCCGATTCCTATACAATAGTACCAGCCCTGTG 391
|||||
Qy 143 ThrSerProThrLysProTyrosThrSerSerProIleLeuSerAspIleLysAlaGlu 162
|||||
Db 392 ACTTCTTCCATTAAATACCTCTCCATCATTTTCTCTACCCAAATATCTCAAGACTGAA 451
|||||
Qy 163 IleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlnAsn 182
|||||
Db 452 ATAAATGTTCCAGAGTCAAGAGTGAATTTGGCCACAGCATCTGCCTGAAGCTAAAT 511
|||||
Qy 183 LysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArgValLeu 202
|||||
Db 512 GAGACTCCCACTGTGAGGAGTTTAAAGAAACAATGAAGAGAGCTGACCCGAGTCTCTG 571
|||||
Qy 203 CysGlyGluGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeuLeuAla 222
|||||
Db 572 TGTGAGGAGGAACAG-----GCCGAGGTGTGCTTGTCTCTGCTGGCC 613
|||||
Qy 223 GlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGluIleSer 242
|||||
Db 614 CAGTCTGAGTGAAGCTCAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 673
|||||
Qy 243 SerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIleLeuAsp 262
|||||
Db 674 AGCAAGATCAGACTTCTGAAAAGGTACCACTGCTGACCTGAGAGAGTGGCGCATCCAGGAC 733
|||||
Qy 263 PheThrGluGlnAspValAlaAlaSerHisGlnSerTyrosGlnLysThrLeuIleAlaLeu 282
|||||
Db 734 ATCTCTGAAGAAGATGCTCAGCAGCCACCAGAGCTATTTCCCGGAAGACCTTGATTCACCTG 793
|||||
Qy 283 ValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrosPheLeuMetAsn 301
|||||
Db 794 GTCACCTCGGGGATCTCTGCTGCCCATCTGATCACCACGCTGATTTCTCTGATGAAC 850
|||||
RESULT 15
HS8812/c
LOCUS
DEFINITION
HS8812 160771 bp DNA linear PRI 22-NOV-1999
Human DNA sequence from clone 8L2 on chromosome 1q32.2-q32.3
Contains CD34 (hematopoietic stem cell surface antigen), CA repeat,
EST, GSS, complete sequence.
ACCESSION AL035091
VERSION AL035091.2 GI:4581417
KEYWORDS HTG; CD34; repeat polymorphism.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 160771)
AUTHORS Direct Submission
TITLE Grafham,D.
JOURNAL Submitted (22-FEB-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Apr 12, 1999 this sequence version replaced gi:4140364.

```

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence is the entire insert of clone 88L2. The true right end of clone 7H11 is at 71906 in this sequence. 8L2 is from the library RPC16 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:ppACA4>.

FEATURES

source

1..160771
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="q32.2-q32.3"
 /clone="XX-D88L2"
 /clone_lib="RPC16"
 42..81
 /note="10 copies 4 mer aaga 83% conserved"
 42..75
 /note="17 copies 2 mer aa 88% conserved"
 89..412
 /note="L1MC5 repeat: matches 7609. .7913 of consensus"
 1027..1154
 /note="L2 repeat: matches 2586. .2744 of consensus"
 1189..1277
 /note="L2 repeat: matches 2616. .2704 of consensus"
 2426..2643
 /note="MER58A repeat: matches 1. .224 of consensus"
 complement(3075..3399)
 /note="match: 252805 STS containing (CA) repeat"
 3121..3176
 /note="28 copies 2 mer gt 100% conserved; differs from 252805"
 3806..3905
 /note="L2 repeat: matches 2600. .2750 of consensus"
 8052..8339
 /note="L2 repeat: matches 1823. .2111 of consensus"
 8406..8600
 /note="MT1A1 repeat: matches 162. .365 of consensus"
 8631..8815
 /note="L2 repeat: matches 2158. .2346 of consensus"
 8843..9132
 /note="AluJb repeat: matches 1. .292 of consensus"
 9295..9363
 /note="L2 repeat: matches 2680. .2749 of consensus"
 9372..9672
 /note="match: GSS AQ075766 clone 2359015"
 9493..9594
 /note="MIR repeat: matches 96. .205 of consensus"
 9619..9640
 /note="11 copies 2 mer tc 100% conserved"
 9671..9835
 /note="MIR repeat: matches 46. .200 of consensus"
 10253..10285
 /note="L2 repeat: matches 2708. .2740 of consensus"
 10284..10406
 /note="MIR repeat: matches 64. .196 of consensus"
 <11347..11584
 /note="match: multiple ESTs; match: AA884400 AA971811"
 11846..11878

/note="MIR repeat: matches 1. .33 of consensus"
 11879..12182
 /note="AluY repeat: matches 1. .303 of consensus"
 12183..12411
 /note="MIR repeat: matches 33. .258 of consensus"
 12443..12491
 /note="L2 repeat: matches 2662. .2710 of consensus"
 12574..12661
 /note="MIR repeat: matches 178. .262 of consensus"
 12756..13049
 /note="AluJo repeat: matches 1. .300 of consensus"
 13231..>13541
 /note="match: STS G28100, G24984"
 13231..>14597
 /note="match: multiple ESTs; match: 3' EST T89361 clone 110082; Paired with EST T84126 matching this clone; match: 5' EST T84126 clone 110082; Paired with EST T89361 matching this clone"
 13672..13875
 /note="MIR repeat: matches 29. .231 of consensus"
 14598..14907
 /note="AluY repeat: matches 1. .310 of consensus"
 14913..15214
 /note="AluJo repeat: matches 7. .294 of consensus"
 15376..15596
 /note="L2 repeat: matches 1263. .1527 of consensus"
 15606..15666
 /note="MLT1F repeat: matches 478. .541 of consensus"
 15667..16045
 /note="L1PA3 repeat: matches 5764. .6146 of consensus"
 16046..16084
 /note="MLT1F repeat: matches 438. .478 of consensus"
 16085..16379
 /note="AluJo repeat: matches 1. .296 of consensus"
 16380..16850
 /note="MLT1F repeat: matches 1. .438 of consensus"
 16854..17208
 /note="L2 repeat: matches 1533. .1930 of consensus"
 17268..17395
 /note="L1MC4 repeat: matches 7854. .7977 of consensus"
 17422..17485
 /note="Charliel repeat: matches 79. .143 of consensus"
 17509..17625
 /note="L1MC4 repeat: matches 7706. .7822 of consensus"
 17627..17677
 /note="L2 repeat: matches 1949. .1995 of consensus"
 17678..17988
 /note="AluY repeat: matches 1. .311 of consensus"
 17989..18227
 /note="L2 repeat: matches 1995. .2257 of consensus"
 18250..18508
 /note="Charliel5 repeat: matches 2315. .2577 of consensus"
 18510..18698
 /note="Charliel5 repeat: matches 16. .196 of consensus"
 18721..18935
 /note="Charlielb repeat: matches 3. .219 of consensus"
 18936..19243
 /note="AluY repeat: matches 1. .307 of consensus"
 19244..19531
 /note="Charlielb repeat: matches 219. .501 of consensus"
 19610..19779
 /note="L1MD repeat: matches 1. .169 of consensus"
 19780..20228
 /note="L1MD repeat: matches 787. .1293 of consensus"
 20299..20606
 /note="L2 repeat: matches 2382. .2710 of consensus"
 20625..20753
 /note="L2 repeat: matches 2179. .2307 of consensus"
 20813..20960
 /note="L2 repeat: matches 2374. .2523 of consensus"
 21149..21292
 /note="MIR repeat: matches 110. .261 of consensus"
 21334..22002


```
misc_feature
<21938..22360
/note="L2 repeat: matches 1947..2724 of consensus"
prim_transcript
<22003..>22108
/note="match: GSS AQ193501 clone 2383C13"
repeat_region
22573..22657
/note="MIR repeat: matches 81..160 of consensus"
misc_feature
complement(24084..25666)
/note="match: STS G06650"
mRNA
complement(join(24102..25486,26245..26409,26724..26776,
27028..27184,35057..35137,36529..36782,37377..37559,
48563..48899))
/gene="CD34"
/note="match: M81104; X60172; match: multiple ESTs; match:
T80274 C18678 AA533071 AA311526 AA311473 AA907722 AA906144
AA064307 A1173145 AA369014 R39493 AA148360 T81557 R72478
R69703 R69740 AA039660 AA639555 R73090 R73955 W92532
R66845 AA054965 AA434483 R69214 R82715 AA188435 AA524276
W7409 H01096 T12007 AA216274 AA022916 AA528783 AA483672
AA640408 W58493 T28504 A1128488 A1149563 A1144193 A1252665
A1017002 A1160509 A1017793 AA906021 A1128776 A1150834
AA434387 AA664247 AA039661 W72884 A1262206"
/evidence=not_experimental
complement(24102..48899)
/gene="CD34"
CDS
complement(join(25301..25486,26245..26409,26724..26776,
27028..27184,35057..35137,36529..36782,37377..37559,
48563..48605))
/gene="CD34"
/note="match: SWISS-PROT:P28906"
/codon_start=1
/evidence=not_experimental
/product="CD34 (hematopoietic stem cell surface antigen)"
/protein_id="CAB36556.1"
/db_xref="GI:4455631"
```

Alignment Scores:

```
Pred. No.: 5,87e-30 Length: 160771
Score: 617.00 Matches: 145
Percent Similarity: 42.3% Conservative: 1
Best Local Similarity: 42.0% Mismatches: 1
Query Match: 40.8% Indels: 198
DB: 9 Gaps: 1
```

US-09-836-602-6 (1-301) x HS88L2 (1-160771)

```
QY 14 LeuProSerGlyPheMetSerLeuAspAsnAsnGlyThrAlaThrProGluLeuProThr 33
::: |||||
Db 37563 ATAGCTTCTGGGTTTCATGAGTCTTGACACACACGGTACTGCTACCCAGAGTTACTACC 37504

QY 34 GlnGlyThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrProSer 53
|||||
Db 37503 CAGGGAACATTTTCAATGTTTCTCAATGTATCTCTACCAAGAACTACAACACCTAGT 37444

QY 54 ThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsn 73
|||||
Db 37443 ACCCTTGGAAGTACCAGCCTGCACCCCTGTGCTCAACATGGCAATGAGGCCACACAAAC 37384

QY 74 IleThr----- 75
|||||
Db 37383 ATCACAGGTAAACACAGCATTTTGTGTAGATCCCGAGAGATGCTGGTGATGCTGGGTA 37324

QY 75 ----- 75

Db 37323 AAGCATTTAGGATGTTTTCAGACCGCTCCCTCTCCACAGAGGAAATATACAAGTCCT 37264

QY 75 ----- 75

Db 37263 AGTATTAAATGACTTGAGTATCATGCTTAGGGTGCCTGAAGTAAAGTTCTAGATAATTCTT 37204

QY 75 ----- 75
```

```
Db 37203 CCTTCAGTGACAGTTTTCTCAGGCCCATGCTTTGGGAGCTGATCCTGATCAGTAATGCCT 37144
QY 75 ----- 75

Db 37143 CCATACCCCTTCTCTCATGCTGAGTCTAGCTTAGTGATAAAATAAGGATCAGAAGACTCA 37084
QY 75 ----- 75

Db 37083 ATAGAGATCTCCGATCTCTTTAAAGGAAAGAGTGGGTACAGGTCCAAACTGGGGGTC 37024
QY 75 ----- 75

Db 37023 TTTGGCTTCTCTAAGGTAGACACGACATCTATTTCAGTTTTCAGGACACCTCTAATACATTC 36964
QY 75 ----- 75

Db 36963 TAGATTTCTAGCTCTGGTTCCAAACAGCTTGAATGAGTTTGGTCAGGGATGGGACACGAAG 36904
QY 75 ----- 75

Db 36903 TAACTGTTAACTCCCCAAACTCCCTTTCTGTGTTGAAGTGCATGCCCATATCATGACCT 36844
QY 75 ----- 75

Db 36843 GAGATTTTGTGTATCTATGACACATGCACATGCACACCCATGTTTGGTCTCTTCCA 36784
QY 76 GluThrThrValLysPheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsn 95
|||||
Db 36783 GAAACGACAGTCAAAATTCACATCTACCTCTGTGATAAACCTCAGTTTATGGAACACAAAC 36724
QY 96 SerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrProAlaAsnVal 115
|||||
Db 36723 TCTTCTGTCAGTCACAGACCTCTGTATATCAGCACAGTGTTCACCACCCAGCCACGTT 36664
QY 116 SerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSer 135
|||||
Db 36663 TCAACTCCAGAGACAAACCTTGAAGCCTAGCCTGTACCTGTGAAATGTTTTCAGACCTTTCA 36604
QY 136 ThrThrSerThrSerLeuAlaThrSerProThrLysProTyrThrSerSerProIle 155
|||||
Db 36603 ACCACTAGCAGCTAGCCTTGCAACATCTCCCACTAAACCCCTATACATCATCTTCTCCTATC 36544
QY 156 LeuSerAspIleLys 160
|||||
Db 36543 CTAAGTGACATCAAG 36529
```

Search completed: October 30, 2002, 09:52:46
Job time : 1587.01 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:10:53 ; Search time 1096.74 seconds
(without alignments)
3888.820 Million cell updates/sec

Title: US-09-836-602-4
Perfect score: 1591
Sequence: 1 MPRGWTALCLLSLLPSPGMS.....YFLMNRSSWSPGTGERLELEP 316

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -Q=/cgn2_1/USPFO_Spool/US09836602/runat_29102002_101144_3812/app_query.fasta_1.1429
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602 -CGCN_1_1_2453 -runat_29102002_101144_3812 -NCPU=6 -ICPU=3
-NO_XUPY -NO_WMAP -LARGEOQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1545	97.1	977	9	AL548050
2	1409	88.6	919	9	AL513812

3	1153	72.5	735	9	AL598821
4	1122	70.5	761	9	AU141253
5	1096	68.9	752	9	AU141111
6	1078	67.8	749	9	AU121093
7	1069.5	67.2	768	9	AU141883
8	1040	65.4	1129	10	BI769224
9	994	62.5	804	10	BG675028
10	991.5	62.3	754	9	AU135182
11	907.5	57.0	1001	10	BG256232
12	792	49.8	717	9	AU121843
13	755.5	47.5	733	9	AU141125
14	714	44.9	825	9	AU139201
15	685.5	43.1	903	10	BI824944
16	672	42.2	780	9	AI806138
17	647	40.7	839	10	BF178239
18	627	39.4	464	10	BM145840
19	627	39.4	551	10	T802174
20	624	39.2	477	10	BM146845
21	578	36.3	473	10	BF830285
22	576.5	36.2	603	9	AA673625
23	572	36.0	486	9	AI893233
24	553	34.8	657	9	AL547643
25	543	34.1	382	10	CI8678
26	533	33.5	464	9	AA792789
27	529	33.2	424	10	W65699
28	470	29.5	997	9	AL547847
29	467	29.4	477	9	AI371138
30	416	26.1	805	10	BG864235
31	415	26.1	397	10	BF993687
32	415	26.1	558	9	AW656254
33	411	25.8	806	10	BE913124
34	365	22.9	223	10	H72215
35	361.5	22.7	675	9	BB665591
36	344.5	21.7	575	9	AA907722
37	338	21.2	297	10	BF471335
38	335	21.1	248	9	AA311473
39	322	20.2	549	10	BE948643
40	316	19.9	253	9	AA311526
41	314	19.7	529	10	BM069265
42	313.5	19.7	551	9	AW910263
43	302	19.0	263	9	AW839633
44	301	18.9	420	9	AI133723
45	297.5	18.7	1081	10	BE916866

ALIGNMENTS

RESULT 1	AL548050	977 bp	linear	EST 16-FEB-2001
LOCUS	AL548050 LTI_NFL006_PL2 Homo sapiens	cdna clone	CSODI034YP22 5	
DEFINITION	prime, mRNA sequence.			
ACCESSION	AL548050			
VERSION	AL548050.1	GI:12882692		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	Li, W.B., Gruber, C., Jessee, J., and Polayes, D.			
TITLE	Full-length cDNA libraries and normalization			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.			
FEATURES	Location/Qualifiers			
source	1..977			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="CSODI034YP22"			
	/clone_lib="LTI_NFL006_PL2"			

/tissue_type="placenta"
/note="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 259 a 272 c 231 g 214 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 9 48e-130 Length: 977
Score: 1545.00 Matches: 311
Percent Similarity: 99.68% Conservative: 0
Best Local Similarity: 99.68% Mismatches: 1
Query Match: 97.11% Indels: 1
DB: 9 Gaps: 0

US-09-836-602-4 (1-316) x AL548050 (1-977)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 40 ATGCCGGGGCTGGACCGCTTTGCTGCTGAGTTTGCTGCCTTCTGGCTTCATGAGT 99
Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 100 CTTGACAAACACGGTACTGCTACCCAGAGTTACTACCCAGGAACATTTTCAATGTT 159
Qy 41 SerThrAsnValSerTyrglnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 160 TCTCAAAATGATCTCTACCAGAACTACACACTAGTACCTTGGAGTACCAGCCTG 219
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVallys 80
Db 220 CACCCTGTGCTCAACATGGCAATGAGGCCAACAAACATCACAGAACGACAGTCAA 279
Qy 81 PheThrSerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValGlnSer 100
Db 280 TTCACATCTACCTCTGTGATACCTCACTGATTATGGAACACAAACTCTCTGTCCAGTCA 339
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 340 CAGACCTCTGTAAACGACAGTCTTACCACCCAGCCACCACTTCACTCCAGAGACA 399
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 400 ACCTTGAAGCCTAGCTGTACCTGGAAATGTTTCAGACCTTTCAACCACTAGCATTGC 459
Qy 141 LeuAlaThrSerProThrLysProTyThrSerSerSerProIleLeuSerAspIleLys 160
Db 460 CTTGCAACATCTCCCACTAAACCCATACATCACTCTCTCTATCCTAAGTGACATCAAG 519
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 520 GCAGAAATCAATGTTTCAAGCATCAGAGAGTGAATGACTCAGGGCATCTGGCTGGAG 579
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 580 CAAAATAAGACCTCCAGCTGTGCGAGTTTAAAGAGACAGGGGAGGGGCTGGCCGA 639
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 640 GTGCTGTGTGGGAGGAGACAGGCTGATGCTGCTGCTGGGCCCAAGGATGCTCCCTGCTC 699
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 700 CTTGCCAGTCTGAGGTGAGGCCCTCAGTGTCTACTGCTGGTCTTGGCCACAGAACAGAA 759
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260

Db 760 ATTTCCAGCAAACTCCAACTTATGAAAAGCACCAATCTGACCTGAAAAGCTGGGGATC 819
Qy 261 LeuaspPheThrGluGlnAspValAlaSerHisGlnSerTygSerGlnLysThrLeuLeu 280
Db 820 CTAGATTTCACTGAGCAAGATGTTTGAAGCCAGCAGAGCTATTCCCAAAGACCCCTGATT 879
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTygPheLeuMet 300
Db 880 GCATGTGTCACTCGGAGCCCTGCTGCTGTCTTGGCAATCACATGGCTATTTCCTGATG 939
Qy 301 AsnArgArgSerTipSerProThrGlyGluArgLeu 312
Db 940 AATGCCGCACGTGGAGCCCMCA-GGAGAAAGGCTG 974

RESULT 2
AL513812
LOCUS
DEFINITION AL513812 LTI_NFL006_PL2 919 bp mRNA linear EST 13-FEB-2001
prime, mRNA sequence.

ACCESSION AL513812
VERSION AL513812.1 GI:12777306
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 919)
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source

1. 919
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CLOBA009ZF10"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 246 a 253 c 213 g 198 t 9 others
ORIGIN

Alignment Scores:

Pred. No.: 1.77e-117 Length: 919
Score: 1409.00 Matches: 286
Percent Similarity: 97.61% Conservative: 0
Best Local Similarity: 97.61% Mismatches: 7
Query Match: 88.56% Indels: 1
DB: 9 Gaps: 0

US-09-836-602-4 (1-316) x AL513812 (1-919)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20

Db 40 ATGCCGGGGCTGGACCGCTTTGCTGCTGAGTTTGCTGCCTTCTGGGTTTCATGAGT 99

Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40

Db 100 CTTGACAAACACGGTACTGCKACCCAGAGTTACTACCCAGGAACATTTTCAATGTT 159


```
Db 628 GAGTGTCTGTGGGAGGAGCAGGCTGATGCTGATGCTGGGCCCGCCAGGTATGCTNCCTGC 687
QY 220 euLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeu 235
|||||
Db 688 TCCTGTCCCACTGTGAGGAGGCCCTCAGTGTCTACTGTGCTGTTG 734
RESULT 4
LOCUS AU141253 761 bp mRNA linear EST 25-OCT-2000
DEFINITION AU141253 THYR01 Homo sapiens cDNA clone THYR01000286 5', mRNA
sequence.
ACCESSION AU141253
VERSION AU141253.1 GI:11002774
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 761)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source Location/Qualifiers
1. .761
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01000286"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/note="Vector: pME18SFL3"
BASE COUNT 197 a 214 c 180 g 167 t 3 others
ORIGIN
Alignment Scores:
Pred. No.: 1,28e-91 Length: 761
Score: 1122.00 Matches: 222
Percent Similarity: 98.68% Conservative: 2
Best Local Similarity: 97.80% Mismatches: 3
Query Match: 70.52% Indels: 0
DB: 9 Gaps: 0
US-09-836-602-4 (1-316) x AU141253 (1-761)
QY 1 MetProArgGlyThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 73 ATCGCGGGGGCTGAGCCGGCTTTGCTGTGAGTTGCTGCTTCTGGTTCATCAGT 132
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 133 CTTGACAAACAGGTACTGTCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 192
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 193 TCTACAAATGTATCTCTACCAAGAACTTACAACACCTAGTACCTCTGGAAGTACCAGCTG 252
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValValys 80
|||||
Db 253 CACCCTGTCTCAACATGGCAATGAGCCCAACACGATCACAGAAACGACAGTCAAA 312
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
```

```
|||||
Db 313 TTCACATCTACCTGTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 372
QY 101 GlnThrSerValIleSerThrValPheThrThrProIleAsnValSerThrProGluThr 120
|||||
Db 373 CAGACCTCTGTATATCAGCAGAGTGTTCACCACCCCAACAGTTTCAACTCCAGAGACA 432
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 433 ACCTTGAAGCTACCTGTCACCTGGAATGTTTCAGACCTTTCAACCACCTAGCAGTAC 492
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLys 160
|||||
Db 493 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTTCAAGTGACATCAAG 552
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
|||||
Db 553 GCAGAAATCAAAATGTTTCAGGCATCAGAGAGTGAATGACTCAGGGCATCTGCCTGGAG 612
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
|||||
Db 613 CAAATAAGACCTCCAGCTGTGCGAGTTTAAAGAAGCAGGAGGAGGCTGCGCCGA 672
QY 201 ValLeuCysGlyGluGlnGlnAlaAspAlaGlyAlaGlnValCysSerSerLeuLeu 220
|||||
Db 673 GTGCTGTGTGGGAGGAGCANGCTGATGCTGCTGGGCCCAAGTATGCTCCTGNTC 732
QY 221 LeuAlaGlnSerGluValArg 227
|||||
Db 733 CTTGCCAGTCTGANGTGAAG 753
RESULT 5
LOCUS AU141111 752 bp mRNA linear EST 25-OCT-2000
DEFINITION AU141111 THYR01 Homo sapiens cDNA clone THYR01000038 5', mRNA
sequence.
ACCESSION AU141111
VERSION AU141111.1 GI:11002632
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 752)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source Location/Qualifiers
1. .752
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01000038"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/note="Vector: pME18SFL3"
BASE COUNT 197 a 208 c 176 g 168 t 3 others
ORIGIN
Alignment Scores:
Pred. No.: 2.84e-89 Length: 752
```

Score: 1096.00 Matches: 218
 Percent Similarity: 98.20% Conservatives: 0
 Best Local Similarity: 98.20% Mismatches: 4
 Query Match: 68.89% Indels: 0
 DB: 9 Gaps: 0

US-09-836-602-4 (1-316) x AU141111 (1-752)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuProSerGlyPheMetSer 20
 DB 65 ATGCCGGGGCTGGACCGCGCTTGTCTGAGTTGCTGCTTCTGGGTTCATGAGT 124
 QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 DB 125 CTTGACAAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 184
 QY 41 SerThrAsnValSerGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
 DB 185 TCTACAAATGATCTCTACCAAGAACTACAAACACTAGTACCTTGGAGTACGAGCTG 244
 QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
 DB 245 CACCTGTGTCTCAACATGCATGAGCCACAAACATCACAGAAACAGAGTCAAA 304
 QY 81 PheThrSerThrSerValIleThrSerValTrGlyAsnThrAsnSerSerValGlnSer 100
 DB 305 TTCACATCTACCTGTGTGATAACCTCAGTTTATGAAACACAAACTCTTCTGTCCAGTCA 364
 QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 DB 365 CAGACCTCTCTAATCAGCAGAGTGTTCACCACCCGACCAACGTTTCAACTCCAGAGACA 424
 QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
 DB 425 ACCTTGAGCCTAGCCGTGCTGCACTGGAAATGTTTCAGACCTTTCAACCATAGCACTAGC 484
 QY 141 LeuAlaThrSerProThrLysProThrThrSerSerSerProIleLeuSerAspIleLys 160
 DB 485 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTTCAAGTACATCAAG 544
 QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
 DB 545 GCAGAAATCAAAATGTTTCAGGCATCAGAGAGTGAATGACTCAGGGCATCTGCTCGAG 604
 QY 181 GlnAsnLysThrSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
 DB 605 CAAAATAGACCTTCAGCTGTGGAGATTTAAGAGACAGGGAGGGCTGGCCCGA 664
 QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
 DB 665 GTGCTGTGTGGGAGGAGCANGCTGATGCTGATGCTGGGCGCCAAAGTATTGCTCTCTC 724
 QY 221 LeuAla 222
 DB 725 CTTGCC 730

RESULT 6
 AU121093
 LOCUS AU121093 HEMBB1 Homo sapiens cDNA clone HEMBB1002065 5', mRNA
 DEFINITION

Accession
 AU121093
 Version
 AU121093.1
 Keywords
 EST.

Source
 human.
 Organism
 Homo sapiens

Reference
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 (bases 1 to 749)

Authors
 Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
 Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and
 Isogai, T.
 HRI human cDNA project

JOURNAL COMMENT

Unpublished (2000)
 Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp

HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.

FEATURES

Location/Qualifiers
 1..749
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HEMBB1002065"
 /clone_lib="HEMBB1"
 /tissue_type="whole embryo, mainly body"
 /dev_stage="embryo, 10 weeks"
 /note="Vector: pME18SFL3"

BASE COUNT 195 a 211 c 173 g 166 t 4 others
 ORIGIN

Alignment Scores:

pred. No.: 1..2e-87 Length: 749
 Score: 1078.00 Matches: 214
 Percent Similarity: 96.83% Conservatives: 0
 Best Local Similarity: 96.83% Mismatches: 7
 Query Match: 67.76% Indels: 0
 DB: 9 Gaps: 0

US-09-836-602-4 (1-316) x AU121093 (1-749)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
 DB 73 ATGCCGGGGCTGGACCGCGCTTGTCTGAGTTGCTGCTTCTGGGTTCATGAGT 132
 QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 DB 133 CTTGACAAACGGTACTGCTTACCCAGAGTACTACCCAGGGAACATTTTCAATGTT 192
 QY 41 SerThrAsnValSerGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
 DB 193 TCTACAAATGATCTCTACCAAGAACTACAACTAGTACCTTGGAGTACGAGCTG 252
 QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
 DB 253 CACCTGTGTCTCAACATGCAATGAGCCACAAACATCACAGAAACGACAGTCAAA 312
 QY 81 PheThrSerThrSerValIleThrSerValTrGlyAsnThrAsnSerSerValGlnSer 100
 DB 313 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTCTGTCTCAGTCA 372
 QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 DB 373 CAGACCTCTGTATCAGCACAGTGTTCACCACCCAGCCCAACGTTTCAACTCCAGAGACA 432
 QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
 DB 433 ACCTTGAAGCCTTACCCGCTCACCTTGGAAATGTTTTCAGACCTTTCAACCATAGCACTAGC 492
 QY 141 LeuAlaThrSerProThrLysProThrThrSerSerSerProIleLeuSerAspIleLys 160
 DB 493 CTTGCAACATCTCCCACTAAACCTTATACATCATCTTCTCTATCTTCAAGTACATCAAG 552
 QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
 DB 553 GCAGAAATCAAAATGTTTCAGGCATCAGAGAGTGAATGACTCANGGCATCTGCTGGAG 612
 QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
 DB 613 CAAAATAGACCTTTCAGCTGTGCGAGTTTAAAGAGGACAGGGGAGGCGCTGGCCCGA 672

```

QY 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlnValCysSerLeuLeu 220
|||||
Db 673 GTGCTGTGGGAGGAACAGCTGATGCTGATGCTGGGGCCCAAGTACTGCTCTCTCC 732

QY 221 Leu 221
|||
Db 733 TTG 735

RESULT 7
AUI41883
LOCUS AUI41883 THYR01 Homo sapiens cDNA clone THYR01001399 5', mRNA
DEFINITION sequence.
ACCESSION AUI41883
VERSION AUI41883.1 GI:11003404
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 768)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
JOURNAL Contact: Takeo Isogai
COMMENT Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1..768
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01001399"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/note="Vector: pME18SFL3"
BASE COUNT 198 a 215 c 181 g 171 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 7 34e-87 Length: 768
Score: 1069.50 Matches: 217
Percent Similarity: 95.61% Conservative: 1
Best Local Similarity: 95.18% Mismatches: 10
Query Match: 67.22% Indels: 1
DB: 9 Gaps: 0

US-09-836-602-4 (1-316) x AUI41883 (1-768)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 72 ATGCCCGGGCTGGACCGCTTTGCTGCTGAGTTTCTGCTCTCTGGTTTCATGAGT 131

QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 132 CTGTCAACAACAGGACTACTCTACCCAGAGTTACCTACCCAGGAGACATTTTCAATGTT 191

QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 192 TCTACAAATGATCTCTACCAAGAACTACACACCTAGTACCTTGGAGTAGCCAGCCTG 251

QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
|||||

```

```

Db 252 CACCCCTGTGTCTCAACATGGCAATGAGGCCAACAAACACACACAGCAGCTCAA 311
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||||
Db 312 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAACTCTTCTGTCCAGTCA 371
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||||
Db 372 CAGACCTCTGTAATCAGCACAGTGTTCACCAACCCAGCCAGCGTTTCAACTCCAGAGACA 431
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 432 ACCTTGAAGCCCTAGCCTGTGCACCTGGAAATGTTTCAGACCTTTCAACCACTAGCACTAG 491
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 492 CTTGCAACATCTCCCACTAAACCTATATACATCATCTTCTCTATCTTAAGTGACATCAAG 551
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
|||||
Db 552 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAAATTGACTCAGGGCATCTGCCTGGAG 611
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlyLeuAlaArg 200
|||||
Db 612 CAAAATAAGACCTTTCAGCTGTGCGGAGTTTAAAGAGACAGGGGAGGCGCTGGCCCGA 671
QY 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlnValCysSerLeuLeu 220
|||||
Db 672 GTGCTGTGGGGAGGAACANGCTGATGCTGATGCTGGGC--CCAGTATGCTTCTCTGCTC 729
QY 221 LeuAlaGlnSerGluValArgPro 228
|||||
Db 730 CTTGCCAATCTGANGTGAAGGCCT 753

RESULT 8
B1769224
LOCUS B1769224
DEFINITION mRNA sequence.
ACCESSION B1769224
VERSION B1769224.1 GI:15760802
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11525 row: e column: 14
High quality sequence stop: 765.
FEATURES
source
1..1129
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5209333"
/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed)

```

upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC Library."

BASE COUNT 292 a 317 c 279 g 241 t
ORIGIN

Alignment Scores:
Pred. No.: 6,01e-84 Length: 1129
Score: 1040.00 Matches: 247
Percent Similarity: 81.70% Conservative: 12
Best Local Similarity: 77.92% Mismatches: 31
Query Match: 65.37% Indels: 27
DB: 10 Gaps: 7

US-09-836-602-4 (1-316) x B1769224 (1-1129)

```
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
|||
Db 55 ATGCGCGGGGCTGGACCGCTTGTGCTGCTGAGTTTCTGCTTCTGGGTTTCATGACT 114
|||
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||
Db 115 CTGCAACAACGGTACTCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 174
|||
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
|||
Db 175 TCTACAAATGATCTTACCAAGAACTACACACCTAGTACCTTGGAGTAGTACCAGCTG 234
|||
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrAsnIleThrGluThrValLys 80
|||
Db 235 CACCCTGTGCTCAACATGGCAATGAGGCCACAAACAACATCACAGAAACGACGTCAAA 294
|||
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
|||
Db 295 TTCACATCTACCTCTGTGATACTCAGTTTATGGAAACACAAACCTTCTGTCTCAGTCA 354
|||
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||
Db 355 CAGACCTCTGTAATCAGCACAGTGTTCACACCCAGCCCAACGTTTCACTTCAGAGACA 414
|||
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||
Db 415 ACCTTGAAGCCTAGCTGTCACTGCAATGTTTTCAGACCTTTCAACACCTAGCAGTACG 474
|||
QY 141 LeuAlaThrSerProThrLysProThrThrSerSerProIleLeuSerAspIleLeu 160
|||
Db 475 CTTGCAACATCTCCCACTAAAGCCCTATACATCATCTTCTCTATCTTAAGTGACATCAA 534
|||
QY 160 sAlaGluIleLysCysSerGlyIle-ArgGluVal-LysLeuThrGlnGlyIleCysLeu 179
|||
Db 535 GCGAGAAATCAATGTTTCAGGATCATCAGAGAGTGCATTTGACTCAGGGCATCTGCCCTG 594
|||
QY 180 GluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla 199
|||
Db 595 GAGCAAAATAAGACCTCCAGCTGTGCGGAGTTTAAGAAGACAGAGGGGAGGGCCCTGGCC 654
|||
QY 200 -ArgValLeuCys--GlyGluGluGlnAlaAspAlaAspAlaGlyAla-Gln--ValC 217
|||
Db 655 CCGAGTGTCTGTGCTGGCGGAGGAGGAGGCTGATGTGATGTGGGGCCCAAGATGTC 714
|||
QY 217 ysSerLeuLeuLeuAlaGlnSerGlu-ValArgPro-GlnCysLeuLeuValLeuAl 236
|||
Db 715 TCCCGTGTCTTGGCCAGCTGACGCGTGAGGCCCTCCAGTGCTACTGCTGGGTCTTGC 774
|||
QY 236 a-----AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisG1 252
|||
Db 775 GCCCCTAGTAAGGGGAATTTGCCCGAGCAACAGTCTCCAGCTTTATGAAAAGACCCCA 834
|||
QY 252 nSerAspLeuLysLysLeu-----GlyIleLeuAspPheThrGluGlnAspValAlase 270
|||
Db 835 ATCTGAGCCTGAACAACAGCTGGGGGAATCCATAGAAATTTGCTGAGCAAGATTGTTGCAC 894
|||
```

```
QY 270 rHisGlnSerTyrGlnLysThrLeuIleAla-----Le 282
|||
Db 895 AAGCATAGC-----AGAACTATTTTCCCAAGAAGACCCCTGATTGGCACCT 945
|||
QY 282 uValThr---SerGlyAlaLeuLeuAlaValLeuGly 293
|||
Db 946 GGAGACCCCTCGGGAAGCCCTGCTGGATGACTTGGG 982
|||
RESULT 9
BG675028 804 bp mRNA linear EST 01-MAY-2001
LOCUS 602621258Fl NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4746591 5',
DEFINITION mRNA sequence.
ACCESSION BG675028
VERSION BG675028.1 GI:13906424
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 804)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue procurement: James Cleaver, M.D.
cDNA library preparation: Life Technologies, Inc.
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Place: LLNL0595 row: d column: 16
High quality sequence stop: 686.
FEATURES
Location/Qualifiers
1..804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4746591"
/cdb="NCI_CGAP_Skn3"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-Sport6; Site: 1: NCI;
Site: 2: Sali; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 219 a 217 c 191 g 177 t
ORIGIN
Alignment Scores:
Pred. No.: 5,35e-80 Length: 804
Score: 994.00 Matches: 229
Percent Similarity: 91.41% Conservative: 5
Best Local Similarity: 89.45% Mismatches: 10
Query Match: 62.48% Indels: 12
DB: 10 Gaps: 2
US-09-836-602-4 (1-316) x BG675028 (1-804)
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
|||
Db 46 ATGCGCGGGGCTGGACCGCTTGTGCTGCTGAGTTTGTGCTTCTGGGTTTCATGAGT 105
|||
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||
Db 106 CTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTTG 165
|||
QY 41 SerThr-AsnValSerTyrGlnGlu---ThrThrThrProSerThrLeuGlySerThrSe 59
|||
Db 166 TCTTACAAAGTGTCTCTTACCAAGAAAGTACCAACGACCTAGGTACCTTGTGTAAGTACCAG 225
|||
QY 59 rLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVa 79
|||
```



```

Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 274 CTTGACAAACAGGTACTGCTACCCAGAGTTACTACCCAGGAACATTTTCAAAATGTT 333
Qy 41 SerThrAsnValSerTyrGlnGlyThrThrProSerThrLeuGlySerThrSerLeu 60
Db 334 TCTACAAATGATCCTACCAAGAAACTACACACCTAGTACCCCTTGGAGTACCAGCCTG 393
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValValys 80
Db 394 CACCCTGTGTCTCAACATGGCAATGAGGCCACAACAACATCACAGAAACGACAGTCAAA 453
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 454 TTCACATCTACTCTGTGATTAACCTCAGTTATGGAACACAAACTCTCTGTCCAGTCA 513
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 514 CAGACCTCTGTAATCAGACACAGTCTTCAACACCCAGCCAGCGTTTCAACTCCAGAGACA 573
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 574 ACCTTGAAGCCTACCCCTGTCACTGGAAATGTTTCAGACCTTTTCAACCACTAGCAGTAC 633
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 634 CTTGCACATCTTCCATCAAAACCTATCA-TCATCTTCTNCTATNCTAAGTGACATCAAG 692
Qy 161 AlaGluLeuLysCysSerGlyIle 168
Db 693 GCAGAAATCAAAATGTTTCANGCATC 716

RESULT 13
AUI41125 733 bp mRNA linear EST 25-OCT-2000
LOCUS AUI41125 THYR01 Homo sapiens cDNA clone THYR0100063 5', mRNA
DEFINITION sequence.
ACCESSION AUI41125
VERSION AUI41125.1 GI:11002646
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
AUTHORS Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source Location/Qualifiers
1..733
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR0100063"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/notes="Vector: pME18SFL3"
BASE COUNT 199 a 212 c 168 g 150 t 4 others
ORIGIN

```

Alignment Scores:

```

Pred. No.: 1,81e-58 Length: 733
Score: 755.50 Matches: 157
Percent Similarity: 94.61% Conservative: 1
Best Local Similarity: 94.01% Mismatches: 8
Query Match: 47.49% Indels: 2
DB: 9 Gaps: 1

```

US-09-836-602-4 (1-316) x AUI41125 (1-733)

```

Qy 1 MetProArgGlyThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 233 ATGCCCGGGGCTGACCGCGCTTTGCTGTGAGTTTGTCTGCTTCTGGGTTTCATGAGT 292
Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 293 CTTGACAAACAGGTACTGCTACCCAGAGTTACTACCCAGGAACATTTTCAAAATGTT 352
Qy 41 SerThrAsnValSerTyrGlnGlyThrThrProSerThrLeuGlySerThrSerLeu 60
Db 353 TCYACAAATGTATCTACCAAGAAACTACACACCTAGTACCTTGGAGTACCAGCCTG 412
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValValys 80
Db 413 CACCCTGTGTCTCAACATGGCAATGAGGCCACAACAACATCACAGAAACGACAGTCAAA 472
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 473 TTCACATCTACTCTGTGATTAACCTCAGTTATGGAACACAAACTCTCTGTCCAGTCA 532
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 533 CAGACCTCTGTAATCAGACACAGTGTTCACCACCCAGCCAGCGTTTCAACTCCAGAGACA 592
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 593 ACCTTGAAGCCTACCCCTGTCACTGGAAATGTTTTCAGACCTTTTCAACCACTAGCAGTACCT 652
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 653 TGCACATCTGCC---ACTAACCCTATCA-TCATCTTCTNCTATCTAAGTGACATCAAG 708
Qy 161 AlaGluLeuLysCysSerGly 167
Db 709 GCAGAAATCAAAATGTTTCAGGC 729

```

RESULT 14

```

AUI39201 825 bp mRNA linear EST 25-OCT-2000
LOCUS AUI39201 PLACE1 Homo sapiens cDNA clone PLACE1010139 5', mRNA
DEFINITION sequence.
ACCESSION AUI39201
VERSION AUI39201.1 GI:11000722
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
AUTHORS Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

```



```
||||| 579 GACATCAGGCCAGAACTCAAATGTCGCAATCAGACAAAGCGAAACTGACTCCGCG 638
QY 175 GlnGlyIleCysLeuGluGlnAsnLysThrSerCysAlaGlu-PheLysLysAspAr 194
Db 639 CCATCGGCCAAGCGCAGCAAACTAAAGACCTCCAGACTGGCGGGAAGCACACCGAAGGACAA 698
QY 194 gGlyGluGly 197
Db 699 GGGCACAGGG 708
```

Search completed: October 30, 2002, 10:47:56
Job time : 1100.74 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 30, 2002, 08:08:13 : Search time 40.5374 Seconds
(without alignments)
1914.781 Million cell updates/sec

Title: US-09-836-602-4
Perfect score: 1591
Sequence: 1 MPRGWTALCLSLLPSEFMS.....YFLMNRSSPTGERLELEP 316

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlip
-Q=/cgn2_1/USPTO.spool/US09836602/runat_29102002_101145_3837/app_query.fasta_1.1429
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602_8CGN1_1_44_@runat_29102002_101145_3837 -NCFU=6 -ICPU=3
-NO_XLPAY -NO_MMAB -LARGEBQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	921	57.9	1065	2 US-08-475-634D-18
2	144	9.1	5163	3 US-08-700-651-1
3	144	9.1	5163	3 US-08-928-361B-4
4	144	9.1	5318	3 US-08-700-651-2
5	144	9.1	5318	3 US-08-928-361B-3
6	137.5	8.6	5511	3 US-08-928-361B-2
7	137.5	8.6	7334	3 US-08-928-361B-1
8	129	8.1	1505	1 US-07-915-246-1
9	128	8.0	390	4 US-09-137-649-7
10	127.5	8.0	3337	1 US-08-072-610-1
11	127.5	8.0	3337	2 US-08-719-822B-1
12	127.5	8.0	3337	4 US-09-092-458-1

13	125.5	7.9	2589	1	US-08-325-267A-3	Sequence 3, Appli
14	125.5	7.9	2685	3	US-08-362-525-21	Sequence 21, Appl
15	125.5	7.9	4614	1	US-08-325-267A-1	Sequence 1, Appli
16	123	7.7	2754	1	US-08-270-076A-10	Sequence 10, Appl
17	121	7.6	2214	6	5258502-1	Patent No. 5258502
18	121	7.6	3168	4	US-09-165-239A-3	Sequence 3, Appli
19	119.5	7.5	688	4	US-08-998-416-915	Sequence 915, App
20	119.5	7.5	1107	2	US-08-991-300-1	Sequence 1, Appli
21	117	7.4	2584	3	US-08-758-662-8	Sequence 8, Appli
22	116	7.3	2793	1	US-08-209-747-1	Sequence 1, Appli
23	116	7.3	2793	1	US-08-458-298-1	Sequence 1, Appli
24	115.5	7.3	2032	4	US-09-241-581B-5	Sequence 5, Appli
25	115.5	7.3	2032	5	PCT-US95-07721-5	Sequence 5, Appli
26	114.5	7.2	2093	1	US-08-287-001A-1	Sequence 1, Appli
27	114.5	7.2	2093	5	PCT-US95-09941-1	Sequence 1, Appli
28	113.5	7.1	2164	4	US-08-760-615-3	Sequence 3, Appli
29	113	7.1	36519	3	US-08-923-137-2	Sequence 2, Appli
30	108.5	6.8	4108	4	US-08-981-729-8	Sequence 8, Appli
31	108.5	6.8	4108	4	US-08-981-446B-1	Sequence 1, Appli
32	108	6.8	2150	2	US-08-861-464-13	Sequence 13, Appl
33	108	6.8	2150	2	US-08-396-001-13	Sequence 13, Appl
34	108	6.8	2150	4	US-08-323-433A-13	Sequence 13, Appl
35	108	6.8	9636	1	US-08-323-170B-1	Sequence 1, Appli
36	108	6.8	9636	4	US-08-954-441-1	Sequence 1, Appli
37	107.5	6.8	3833	1	US-08-917-320-18	Sequence 18, Appl
38	107.5	6.8	3833	5	PCT-US95-04611A-18	Sequence 18, Appl
39	107.5	6.8	5931	3	US-08-783-774-1	Sequence 1, Appli
40	106	6.7	10803	3	US-09-080-044-1	Sequence 1, Appli
41	106	6.7	19227	3	US-09-090-793-13	Sequence 13, Appl
42	106	6.7	40138	3	US-09-090-793-12	Sequence 12, Appl
43	105	6.6	3141	2	US-08-658-665-66	Sequence 66, Appl
44	105	6.6	3141	4	US-08-796-101-30	Sequence 30, Appl
45	105	6.6	3141	4	US-09-085-273-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1
US-08-475-634D-18
; Sequence 18, Application US/08475634D
; Patent No. 5962844
; GENERAL INFORMATION:
; APPLICANT: HAWLEY, Robert J.
; APPLICANT: MONROY, Rodney L.
; TITLE OF INVENTION: Antibodies to Porcine CD34 Positive
; TITLE OF INVENTION: and Uses Therefor
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS 6.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,634D
; FILING DATE: June 7, 1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 61750-141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 973-994-1700
; TELEFAX: 973-994-1744
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:


```
Qy 204 GlyGluGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeuLeuAlaGln 223
Db 3334 ATCGAAAGATGACGCTATTGAAATGGATGCATTTACAAATGATCCAAATGATGAC 3393
Qy 224 SerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGluIle----- 241
Db 3394 ACGCATGTCCTTCAGATTTAAGGTTAAAGATGATAGGAACACTATTTCAGTAAGATGC 3453
Qy 242 -----SerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeu 258
Db 3454 GGAAGAGTGCGAGTAACATCGAGTCCAGATAGAACT----- 3492
Qy 259 GlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThr 278
Db 3493 -----TTGGATTTCACAATTCCTCCAGTAGCTGGCCATACACTGTTCA----- 3537
Qy 279 LeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaVal 291
Db 3538 ATAATAGTTGGTGTGAGCGCGGTGGAAAAATTCACGTA 3576
RESULT 8
US-07-915-246-1/c
; Sequence 1, Application US/07915246
; Patent No. 5401836
; GENERAL INFORMATION:
; APPLICANT: Baszczynski, Chris L.
; APPLICANT: Fallis, Lynne
; APPLICANT: Bellmare, Guy
; APPLICANT: Bolvin, Rodolphe
; TITLE OF INVENTION: A BRASSICA REGULATORY SEQUENCE FOR
; TITLE OF INVENTION: ROOT-SPECIFIC OR ROOT-ABUNDANT GENE EXPRESSION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McAndrews, Held, and Malloy
; STREET: 500 W. Madison St. 34th Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/915,246
; FILING DATE: 19920716
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 91 P 1125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 707-8889
; TELEFAX: 312 707-9155
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1505 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Brassica napus
; STRAIN: Westar
; DEVELOPMENTAL STAGE: Somatic
; TISSUE TYPE: Root
US-07-915-246-1
```

Alignment Scores:

Pred. No.: 0.000333 Length: 1505

```
Score: 129.00 Matches: 47
Percent Similarity: 48.78% Conservative: 33
Best Local Similarity: 28.66% Mismatches: 70
Query Match: 8.11% Indels: 14
DB: 1 Gaps: 5
US-09-836-602-4 (1-316) x US-07-915-246-1 (1-1505)
Qy 12 SerLeuLeuProSerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeu 31
Db 1370 AGTGCATACCTCCCTCCCTCTCTCCGCCACCTCCACCCACCCACCGCTCTCTCC 1311
Qy 32 ProThrGlnGlyThrPheSer-----AsnValSerThrAsnValSerTyr 46
Db 1310 ACCACCGTGCTCTCTCTCTCCACCCACCCACCCATATCCCTCTCCAGCTCCCGGC 1251
Qy 47 GlnGluThrThrProSerThrLeuGlySerThrSerLeuHisProValSerGlnHis 66
Db 1250 ACCATAACACCCACCGTGAGCTCCACCGGGGCCAGAACCTCTCTCCACCGCTCTCTT 1191
Qy 67 GlyAsnGluAlaThrThrAsnIle-----ThrGluThrThrValLysPheThrSer 83
Db 1190 TCCCCCGCTCCACCCACCTCCATGTCCACCTGCACCGCCACCTCCATATCTCTCCAGC 1131
Qy 84 ThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSerGlnThrSer 103
Db 1130 ACCAGCGCTCTCTCCACCCACCGTATCCACCCACCGTGACCTCCACCTCTCTCCACCGC 1071
Qy 104 ValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThrThrLys 123
Db 1070 TCCATTTCCACCGCTCCACCGACC-----TCCATGTCCACCTCCACCGCCACCC- 1023
Qy 124 ProSerLeuSerProGlyAsnValSerAspLeuSerThrThr-----SerThrSerLeuAla 142
Db 1022 ---TCCATATCTCTCTCCACCGACCCACCTCTCCACCCACCGTATCCACCATCCAGCGC 966
Qy 143 ThrSerProThrLysProTyrThrSerSerSerProTyrLeuSerAspIleLysAlaGlu 162
Db 965 ACGCGCAGCTCTCTCCACCCACCCACCTCTCTCCCTTCTCTCCACCATGCTCTCTCCAGC 906
Qy 163 IleLysCysSer 166
Db 905 TCCTCCCGCTCC 894
RESULT 9
US-09-197-649-7
; Sequence 7, Application US/09197649
; Patent No. 6194550
; GENERAL INFORMATION:
; APPLICANT: Gold, Larry
; APPLICANT: Tuerk, Craig
; APPLICANT: Pribnow, David
; APPLICANT: Smith, Jonathan D.
; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
; FILE REFERENCE: NEX02/C1-CON
; CURRENT APPLICATION NUMBER: US/09/197,649
; CURRENT FILING DATE: 1998-11-23
; EARLIER APPLICATION NUMBER: 07/829,461
; EARLIER FILING DATE: 1992-01-31
; EARLIER APPLICATION NUMBER: 07/739,055
; EARLIER FILING DATE: 1991-08-01
; EARLIER APPLICATION NUMBER: 07/561,968
; EARLIER FILING DATE: 1990-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence
; OTHER INFORMATION: having a 120 repeat of ACG flanked by fixed
; OTHER INFORMATION: fragments having NcoI restriction sites.
```


;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Darby and Darby
;; STREET: 805 Third Ave.
;; CITY: New York
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10022-7513
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; FILING DATE: 09/30/96
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Gogoris, Adda
;; REGISTRATION NUMBER: 29,714
;; REFERENCE/DOCKET NUMBER: 5986/17686US2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212)527-7700
;; TELEFAX: (212)753-6237
;; TELEX: 236687
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3337 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Plasmodium vivax
;; IMMEDIATE SOURCE:
;; CLONE: PVMB3.3.1
;; US-08-719-822B-1

Alignment Scores:
Pred. No.: 0.00191 Length: 3337
Score: 127.50 Matches: 53
Percent Similarity: 41.88% Conservative: 27
Best Local Similarity: 27.75% Mismatches: 96
Query Match: 8.01% Indels: 15
Gaps: 6

US-09-836-602-4 (1-316) x US-08-719-822B-1 (1-3337)

Qy 5 TrpThrAlaLeuCysLeuLeuSerLeuLeuPro-----SerGlyPheMetSerLeu 21
Db 2527 TGGTAGTCGATGCGATGATAGATAAAATCCAGAGCCCTAATTCGCTCTTTCTTTGGTTG 2468
Qy 22 AspAsn-----AsnGlyThrAlaThrProGluLeuProThrGlnGlyThr 36
Db 2467 TCTAATGTATATCTCTTCGCGATGGTATTACTAATGTGTAATACATCTCTCTCTACT 2408
Qy 37 PheSerAsnValSerThrAsnValSerThrGlnGluThrThrProSerThrLeuGly 56
Db 2407 GGTCTTCCT 2348
Qy 57 SerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGlu 76
Db 2347 GGTACT 2288
Qy 77 ThrThrValLysPheThrSerThrSerValIleThrSerValTyGlyAsnThrAsnSer 96
Db 2287 ACTTCTCTCTGCTACCTCTCTCCACTTCTCTGCTACCTCTCTCTCTCTCTCTCTCTCT 2228
Qy 97 SerValGlnSerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSer 116
Db 2227 TCCACTTCTCTGTTGACTCTCT-----TCCACTTCTCTCTCTCTCTCTCTCTCTCTCT 2174

Qy 117 ThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThr 136
Db 2173 ACTTCTCTGTTGATACCTCTTCACACTTCTTCTACTTCTGCGAGGTACCTCT-----TCCACT 2120
Qy 137 ThrSerThrSerLeuAlaThrSerProThr---LysProTyThrSerSerSerProIle 155
Db 2119 TCTTCTACTTCTGCGAGGTACCTCTTCCACTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2063
Qy 156 LeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGln 175
Db 2062 ---TCTACGGTTTCTTCTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2006
Qy 176 GlyIleCysLeuGluGlnAsnLysThrSerSer 186
Db 2005 GGAGTTGCCTCTAATTTCTTCTTCTGCGACTTCT 1973
RESULT 12
US-09-092-458-1/c
; Sequence 1, Application US/09092458
; Patent No. 6231861
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,458
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/719,821
; FILING DATE: 09/30/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/17686US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3337 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: PVMB3.3.1
; US-09-092-458-1
Alignment Scores:
Pred. No.: 0.00191 Length: 3337
Score: 127.50 Matches: 53
Percent Similarity: 41.88% Conservative: 27
Best Local Similarity: 27.75% Mismatches: 96
Query Match: 8.01% Indels: 15


```
Db 1705 ACATATTTCTTCTTCATCATTACCACTTGTACCAGTGCACACACAGCCAGGAATGCT 1764
QY 196 GluGlyLeuAlaArgValLeuCysGlyGluGluGluAlaAspAla----- 210
Db 1765 TCTTCATTAC--CACCTGTACCACTACAAAACAGCAGCAAAACCACTTTGGTTACCG 1822
QY 211 ---AspAlaGlyAlaGlnValCysSerLeuLeuLeuAlaGlnSerGluValArgProGln 229
Db 1823 TGACATCTCGCAGTCTCATGTGTCGACATGATCCATCTCCCTGGGATTTTCCACAG 1882
QY 230 CysLeuLeuValLeuAlaAsnArgThrGluIle---SerSerLysLeuGlnLeuMet 248
Db 1883 CTACTGTTACTGTAGCGCGTCACACACAGATATCCACATGGTGCCTATTCTTACTA 1942
QY 249 LysLysHisGlnSerAspLeuLysLysLeuGlyIleLeuAspPheThr----- 264
Db 1943 CAGAGACAAACAGC-----AAACCAA-GGGACACACAGCAACACACAGAAACAA 1995
QY 265 -----GluGlnAspValAlaSerHisGlnSer 273
Db 1996 AAACAAACCCAGGTAGTACATTTCTTCTGTGAATCTGACGTATGCTCT----- 2046
QY 274 TyrSerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGly 293
Db 2047 -----AAGACTGCTTCTCCAGCCATGTATCTACAAGCACTGCTACTATTAAACGC 2097
QY 294 IleThrGlyTyrPheLeuMetAsnArgArgSerTrpSerPro 307
Db 2098 GTTACTACAGATAC-----ACAACATGGTGTCT 2127

RESULT 15
US-08-325-267A-1
; Sequence 1, Application US/08325267A
; Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI
; APPLICANT: TAKATA, YOSHIHIRO
; APPLICANT: OGAWA, MASAHIRO
; APPLICANT: PENTTILA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, SIRKKA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; TITLE OF INVENTION: CONTAINING THEM
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSER: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,267A
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP PCT/JP94/00290
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 38871/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
```

```
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4614 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; STRAIN: ABX1-ID
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4614
; US-08-325-267A-1

Alignment Scores:
Pred. No.: 0.00549 Length: 4614
Score: 125.50 Matches: 85
Percent Similarity: 41.32% Conservative: 53
Best Local Similarity: 25.45% Mismatches: 109
Query Match: 7.89% Indels: 89
DB: 1 Gaps: 16

US-09-836-602-4 (1-316) x US-08-325-267A-1 (1-4614)

QY 26 ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnValSer 45
Db 3124 ACAACCAACCACTGAA--CCATGGACTGGTACTTTTACTTCGACTTCCCACTGAATGTCT 3180
QY 46 -----TyrGlnGlyThr-----ThrPro 52
Db 3181 ACTGTCACCTGGAAACCAATGGCTTGCACACTGATGAACCTGTCATTTGTGCAAACTCCA 3240
QY 53 SerThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThr 72
Db 3241 ACTACTGCCATCTCATCCAGTTGTCA-----TCATCATCTTCAGACAAATCACCAGC 3294
QY 73 AsnIleThrGluThrThrValLysPheThrSerThrSerValIleThrSerValThrGly 92
Db 3295 TCTATCACG-----TCTTCGCTGCCAATATTATCCCACTTCTATCTCTATCTCT 3336
QY 93 AsnThrAsnSerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrPro 112
Db 3337 AGCAATGGAACTTCTGTGATTTCT---TCCTCAGTAATTTCT----- 3375
QY 113 AlaAsnValSerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSer 132
Db 3376 TCCTCAGTCACTTCTTCTCTATTCACTTCTTCTCCAGTCATTTCTTCTCAGTCATTTCT 3435
QY 133 AspLeuSerThrThrSerThrSerLeu-----AlaThrSerPro 145
Db 3436 TCTTCTACAAACACCTCCACTTCTATATTTCTGAATCATCTAAATCATCCGTCATTTCCA 3495
QY 146 ThrLysProTyrThrSerSerSerProIleLeuSerAspIleLysAlaGluIleLysCys 165
Db 3496 ACCAGTAGTTCCACCTCTGGTTCT-----TCTGACGCGGAAACG 3534
QY 166 SerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlnAsnLysThrSer 185
Db 3535 AGTTCAGCTGGTTCTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3594
QY 186 -----SerCysAlaGluPheLysLysAspArgGly 195
Db 3595 ACATATTTCTTCTCATCATTTACCCTTGTATCCAGTGCACAAACAGCCAGGAATGCT 3654
QY 196 GluGlyLeuAlaArgValLeuCysGlyGluGluAlaAspAla----- 210
Db 3655 TCTTCTATTAC--CACCTGTCTACCACTACAAAACAGCAGCGAAGCAAACTTTGGTTACCG 3712
QY 211 ---AspAlaGlyAlaGlnValCysSerLeuLeuAlaGlnSerGluValArgProGln 229
Db 3713 TGACATCTCGCAGTCTCATGTGTCACATGATCCATCTCCCTCGCATTTGTTTCCACAG 3772
```


GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:04:33 ; Search time 143.956 Seconds
(without alignments)
3768.837 Million cell updates/sec

Title: US-09-836-602-4

Perfect score: 1591

Sequence: 1 MPRGTALCLLSLPSGMS.....YFLMRRNSWSPGTERLELEP 316

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09836602/runat_29102002_101143_3758/app_query.fasta_1.1429
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US09836602 @CGN_1.1_312 @runat_29102002_101143_3758 -NCFU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPLOG -DEV_FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DLEOP=6 -DELEXT=7

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1591	100.0	951	24	Human CD34 surface
2	1574	98.9	1122	24	Human full-length
3	1574	98.9	2453	22	Nucleotide sequenc
4	1574	98.9	2615	21	Human CD-34 polynu
5	1574	98.9	2615	21	Human adenosine re
6	1574	98.9	3490	21	Human CD-34 polynu
7	1574	98.9	3490	21	Human adenosine re
8	1509	94.8	906	24	Human CD34 surface
9	921	57.9	1065	18	Porcine CD34 from
10	508	31.9	615	21	Human secreted pro
c 11	311	19.5	233	22	Human breast cell
c 12	311	19.5	233	22	Human foetal liver
c 13	311	19.5	233	22	Probe #14530 for g
c 14	311	19.5	233	22	Human brain expres
c 15	311	19.5	233	22	Human bone marrow
c 16	311	19.5	233	22	Probe #13951 for g
c 17	311	19.5	233	22	Probe #18003 used
c 18	311	19.5	233	22	Probe #9593 used t
c 19	265	16.7	209	22	Human breast cell
c 20	265	16.7	209	22	Human foetal liver
c 21	265	16.7	209	22	Probe #14140 for g
c 22	265	16.7	209	22	Human brain expres
c 23	265	16.7	209	22	Human bone marrow
c 24	265	16.7	209	22	Probe #13530 for g
c 25	265	16.7	209	22	Probe #17596 used
c 26	265	16.7	209	22	Probe #9203 used t
c 27	233	14.6	235	22	Novel human diagno
c 28	170.5	10.7	406	22	Human breast cell
c 29	170.5	10.7	406	22	Human foetal liver
c 30	170.5	10.7	406	22	Probe #4241 for ge
c 31	170.5	10.7	406	22	Human brain expres
c 32	170.5	10.7	406	22	Human bone marrow
c 33	170.5	10.7	406	22	Probe #4329 for ge
c 34	170.5	10.7	406	22	Probe #4456 used t
c 35	159	10.0	13154	20	Probe #4210 used t
c 36	159	10.0	13154	20	Enterococcus faeco
c 37	150.5	9.5	2336	23	Drosophila melanog
c 38	148	9.3	40875	18	Insert from cosmid
c 39	147.5	9.3	7720	21	Genomic DNA encodi
c 40	146	9.2	4116	23	Drosophila melanog
c 41	146	9.2	6116	23	Drosophila melanog
c 42	145.5	9.1	5162	23	Drosophila melanog
c 43	144.5	9.1	8298	22	Human immune/haema
c 44	144	9.1	5163	19	Cryptosporidium pa
c 45	144	9.1	5163	21	ORF encoding a por

ALIGNMENTS

RESULT 1
AAD22662
ID AAD22662 standard; DNA; 951 BP.
XX AAD22662;
AC AAD22662;
XX
XX 26-FEB-2002 (first entry)
DT
XX
DE Human CD34 surface antigen truncated variant (tCD34) encoding DNA.
XX
XX Human; surface marker; surface antigen; T lymphocyte; gene therapy;
KW CD34 truncated variant; tCD34; ds.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT CDS 1..951

XX	SQ	Sequence	2463 BP; 589 A; 709 C; 580 G; 585 T; 0 other;
Alignment Scores:			
Pred. No.:	4, 89e-115	Length:	2463
Score:	1574.00	Matches:	313
Percent Similarity:	99.37%	Conservatives:	1
Best Local Similarity:	99.05%	Mismatches:	2
Query Match:	98.93%	Indels:	0
DB:	22	Gaps:	0
US-09-836-602-4 (1-316) x AAH75132 (1-2463)			
Qy	1	MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer	20
Db	127	ATCCGCGGGCTGGACCGCGTTTGCTTGCCTGAGTTGCTTGCCTCTCGGGTTCATGAGT	186
Qy	21	LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
Db	187	CTTGACAAACACGGTACTGCTACCCGAGAGTTTACCTACCCAGGGAAACATTTTCAAAATGTT	246
Qy	41	SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu	60
Db	247	TCTACAAATGATCTCTACCAAGAACTACACACACTAGTACCTTTGGAAGTACCGAGCTG	306
Qy	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys	80
Db	307	CACCCCTGTCTCAACATGGCAATGAGGCCACACAAACATCATCAGAAACGACAGTCAAA	366
Qy	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer	100
Db	367	TTCCATCTACCTCTCTGATACCTCAGTTTATGGAACACAAACTCTTCTGTCAGTCA	426
Qy	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	427	CAGACCTCTGTAATCAGCAGAGTGTTCACCCGCCAGCAACGTTTCAACTCCAGAGACA	486
Qy	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer	140
Db	487	ACCTTGAAGCCTTAGCCTGTACCTGGAATGTTTCAGACCTTTTCAACCACATAGCAGCTAGC	546
Qy	141	LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys	160
Db	547	CTTGCACATCTCCCACTAAACCCATACATCATCTTCTCCTATCCTAAAGTGACATCAAG	606
Qy	161	AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu	180
Db	607	GCAGAAATCAAATGTTTCAGGCGATCAGAGAAGTGAATTTGACTCAGGGCATCTGCGCTGGAG	666
Qy	181	GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg	200
Db	667	CAAAATGAAGACCTCCAGCTGTGCGGAGTTTAAAGAAGACAGGGAGAGGGCTTGCCCGGA	726
Qy	201	ValLeuCysGlyGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeu	220
Db	727	GTGCTGTGTGGGAGGAGCAGCGCTGATGCTGATGCTGGGGCCAGGTATGCTCCTCGTCT	786
Qy	221	LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu	240
Db	787	CTTTGCCCGAGTCTGAGGTGAGGCGCTCAGTGTCTACTGCTGCTTGGCCAAACAGAACAA	846
Qy	241	IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle	260
Db	847	ATTTCCAGCAACTCCAACTTATGAANAAGCACCAACTCTGACCTTGAANAAGCTGGGATC	906
Qy	261	LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle	280
Db	907	CTAGATTTTCACTGAGCAAGATGTTCAAGCCCAACAGACCTATTTCCTCAAAAGACCTTGAT	966
Qy	281	AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet	300
Db	967	GCATGGTCACCTCGGAGCCCTGCTGGCTGTCTTTGGGCATCACTGGGTATTTCCCTGATG	1026

CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.

XX SQ Sequence 2615 BP; 616 A; 763 C; 627 G; 609 T; 0 other;

Alignment Scores:

Pred. No.: 5.28e-115 Length: 2615
Score: 1574.00 Matches: 313
Percent Similarity: 99.37% Conservative: 1
Best Local Similarity: 99.05% Mismatches: 2
Query Match: 98.93% Indels: 0
DB: 21 Gaps: 0

US-09-836-602-4 (1-316) x AAF20898 (1-2615)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 294 ATGCCGGGGGCTGGACGGGCTTGTGCTGAGTTGCTGCTTGGGTTTCATGAGT 353
Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 354 CTTGACACACAGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAAAATGTT 413
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 414 TCTACAATGTATCTACCAAGAATCTACACACCTAGTACCTTGGAGTACACGCGTG 473
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 474 CACCCTGTGTCTCAACATGGCAATGAGGCCACAAACATCAGAGAACGACAGTCAAA 533
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 534 TTCACATCTACCTGTGATAACCTCAGTTTATGGAACACAACTCTCTGTCCAGTCA 593
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 594 CAGACCTCTGTAAACACACAGTGTTCACACCCACCCAGCAACGTTTCAACTCCAGACA 653
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 654 ACCTTGAAGCCTAGCCTGTCACTGGAAATGTTTTCAGACCTTTCAACCACTAGCAGT 713
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 714 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCCTATCTTAAGTGACATCAAG 773
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 774 GCAGAAATCAATGTTTCAAGGATCAGAGAAGTGAATTTGACTCAGGCGATCTGCTGGAG 833
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 834 CAAAATAAGACCTCCAGCTGTGCGAGTTTAAAGAGGACAGGGGAGGCGCTGCCCGGA 893
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlnValCysSerLeuLeu 220
Db 894 GTGCTGTGTGGGGAGGAGCAGGCTGATGCTGATGCTGGGGCCAGGTATGCTCCGTGCTC 953
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 954 CTTGCCCAAGTGTAGGTGAGGCGCTCAGTGTCTACTGTGCTTGTGGCCACAGACAGAA 1013
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 1014 ATTTCCAGCAAACTCCAACTTATGAAAAGACCACTGACCTGAAAACCTGGGGATC 1073
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerThrLysThrLeuIle 280
Db 1074 CTAGATTTCTACTGACGAAGATGTTGCAAGCCACAGAGCTATTCCCAAGACCTGATT 1133

Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 1134 GCACCTGTGCTACCTCGGAGGCGCTGCTGCTGCTTGGCATCACTGCTATTTCTGATG 1193
Qy 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro 316
Db 1194 AATCGCGCGACCTGGAGCCGACAGGAGAAAGGCTGGGCGAGACCT 1241

RESULT 5

AAA34776
ID AAA34776 standard; DNA; 2615 BP.

AC AAA34776;

XX 28-JUL-2000 (first entry)

XX Human adenosine receptor related polynucleotide SEQ ID NO:2465.

XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

OS Homo sapiens.

XX WO200009525-A2.

PN 24-FEB-2000.

XX 03-AUG-1999; 99WO-US17712.

XX 03-AUG-1998; 98US-0095212.

XX (UYEC-) UNIV EAST CAROLINA.

PI Nyce JW;

XX WPI; 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -

XX Disclosure; Page 602; 1343pp; English.

XX The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 18%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, and then the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present

CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.

XX
SQ Sequence 2615 BP; 616 A; 763 C; 627 G; 609 T; 0 other;

Alignment Scores:

Pred. No.: 5,28e-115 Length: 2615
Score: 1574.00 Matches: 313
Percent Similarity: 99.37% Conservatism: 1
Best Local Similarity: 99.05% Mismatches: 2
Query Match: 98.93% Indels: 0
DB: 21 Gaps: 0

US-09-836-602-4 (1-316) x AAA34776 (1-2615)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
DB 294 ATGCCCGGGGCTGGACCGGCTTCTGCTGAGTTTCTGCTTCTGGGTTCATGAGT 353
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
DB 354 CTTGACAAACAGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 413
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
DB 414 TCTACAAATGTATCCTTACCAGAAACTACAAACACCTAGTACCTTGGAAAGTACCAGCCTG 473
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
DB 474 CACCCTGTCTCAACATGGCAATGAGGCCACACAAACATCACAGAAACGAGTCAAA 533
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
DB 534 TTCACATCTACCTCTCTGATACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 593
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
DB 594 CAGACCTCTGTATATCAGCACAGTGTTCACCACCCCAAGCAAGCTTTTCAACTCCAGAGACA 653
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrThrSer 140
DB 654 ACCTTGAAGCCTAGCCTGTCCCTGGAAATGTTTTCAGACCTTTCACCACTAGCAGTACG 713
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
DB 714 CTTGCAACATCTCCCACTAAACCTTATACATCATCTTCTCCATCTTAAAGTACATCAAG 773
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
DB 774 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAAATTTGACTCAGGGCATCTGCCTGGAG 833
QY 181 GlnAspLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluLeuAlaArg 200
DB 834 CAATAATACACCTCCACCTGCTGGAGTTTAGAAGAGCAGGGAGAGGCTGGCCCCA 893
QY 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
DB 894 GTCTGTGTGGGAGGAGCAGCTGATGTGTGTGTGGGCCAGGTATGCTCCCTGCTC 953
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
DB 954 CTTGCCACGTCTGAGGTGAGGCTCAGTGTCTACTGTCTGTGGCCAAACAGAGAA 1013
QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
DB 1014 ATTTCAGCAAACTCAACTTATGAAAAGACCAACTATCGACTGAAAAGCTGGGGATC 1073
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
DB 1074 CTAGATTTTCACTGAGCAAGATGTTGCAAGCCACAGAGTATATCCAAAGACCTGATT 1133
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300

DB 1134 GCACCTGGTCCACCTCGGAGCCCTGCTGGCTGTCTGGGCATCACCTGGCTATTTCCTGATG 1193
QY 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro 316
DB 1194 AATCGCCGAGCTGGAGCCCCACAGGAGAGAGCTGGCGAAGACCCT 1241
RESULT 6
AAF20899
ID AAF20899 standard; DNA; 3490 BP.
XX AAF20899;
AC AAF20899;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human CD-34 polynucleotide fragment #2466.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX (NYCE/) NYCE J W.
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
CC Low adenosine (A) content antisense oligonucleotides which do not
CC trigger adenosine receptors during metabolism, useful e.g. for treating
CC cancers and respiratory obstructions -
XX
XX Disclosure: Page 291-292; 1592pp; English.
CC
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome

CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 3490 BP; 880 A; 966 C; 872 G; 772 T; 0 other;

Alignment Scores:
Pred. No.: 7.68e-115 Length: 3490
Score: 1574.00 Matches: 313
Percent Similarity: 99.37% Conservative: 1
Best Local Similarity: 99.05% Mismatches: 2
Query Match: 98.93% Indels: 0
DB: 21 Gaps: 0

US-09-836-602-4 (1-316) x AAF20899 (1-3490)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 1169 ATGCCGGGGCTGACCGCGCTTGTCTGCTGAGTTTGGCTTCTGGGTTTCATGAGT 1228
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 1229 CTTGACAACAACGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAATGTT 1288
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 1289 TCTCAAAATGTATCTACCAAGAACTACAACACTAGTACCTTGGAAAGTACCAGCCTG 1348
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
Db 1349 CACCTGTGTCTCAACATGGCAATGAGCCACAAACATCACAGAAACGACAGTCAAA 1408
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 1409 TTCACATCTACTCTGTGATAACCTCAGTTATGGAACACAAACTCTTCTGTCCAGTCA 1468
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 1469 CAGACCTCTGTAATCAGCACAGTGTTCACCAACCCAGCCAACTTCAACTCCAGAGACA 1528
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrSerThrSer 140
Db 1529 ACCTTGAGCCTAGCCTGTACCTGGAATGTTTCAGACCTTCAACACCTAGCAGCTAGC 1588
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 1589 CTTGCAACATCTCCACATAAACCCCTATACATCATCTTCTCTATCTTAAGTGACATCAAG 1648
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 1649 GCAGAAATCAAAATGTTACGAGCATCAGAGAAGTGAATTTGACTCAGGGCATCTGCTGGAG 1708
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 1709 CAAATAAGACTCCAGCTGTGCGAGTTTAAAGAGACAGGGAGAGGGCTGGCCCGA 1768
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 1769 GTGCTGTGTGGGAGGACGAGCTGATGCTGATGCTGGGGCCAGGTATGCTCCCTGCTC 1828
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu 240
Db 1829 CTTGCCAGCTGAGGTGAGGCCTCAGTGTCTACTGTGCTGTTGGCCAAACAGACAGAA 1888
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 1889 ATTTCCAGCAACCTCCACTTATGAAAGACCACTATGACCTGAAAGAGCTGGGGATC 1948
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280

Db 1949 CTAGATTTCCTACTGAGCAAGATGTTTCAAGCCACACAGAGCTATTCCTCAAAAGACCTGATT 2008
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 2009 GCATGTGTCACCTCGGAGCCCTGCTGGCTGCTTGGGCATCATCTGCTATTTCTCTGATG 2068
Qy 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuLeuGluLeuPro 316
Db 2069 AATGCCGACACTGGAGCCGCCACAGGAGAAAGGCTGGCGGAGACCT 2116

RESULT 7
AAA34777
ID AAA34777 standard; DNA; 3490 BP.
XX
AC AAA34777;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2466.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.

XX WO200009525-A2.
XX 24-FEB-2000.
XX 03-AUG-1999; 99WO-US17712.
XX 03-AUG-1998; 98US-0095212.
XX (UYEC-) UNIV EAST CAROLINA.
XX Nyce JW;
XX WPI; 2000-205971/18.
XX
XX New antisense oligonucleotides useful for treating e.g. pulmonary
XX vasoconstriction, inflammation, allergies, asthma, hypertension,
XX bronchitis, emphysema, respiratory distress syndrome, ischemia or
XX cancers -
XX
XX Disclosure: Page 603; 1343pp; English.

XX The present invention describes a new composition comprising an
XX antisense oligonucleotide (ON) with low adenosine (up to 15%), which
XX targets nucleic acids involved in bronchoconstriction, allergies, and/or
XX inflammation. The ON can have antiinflammatory, antiallergic,
XX antiasthmatic, cytostatic and analgesic activities. The compositions are
XX useful for the treatment of diseases associated with inflammation,
XX impaired airways, including lung disease and diseases whose secondary
XX effects afflict the lungs of a subject. They can be used for treating
XX e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
XX asthma, impeded respiration, respiratory distress syndrome, pain, cystic
XX fibrosis, pulmonary hypertension, emphysema, chronic obstructive
XX pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
XX carcinomas, and cancers which may metastasize to the lungs, including
XX breast and prostate cancer. The reduction of the adenosine content of
XX the ONs reduces side effects. The A-containing ONs break down with the
XX release of deoxyadenosine which activates adenosine receptors causing
XX bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
XX nucleotide sequences given in the sequence listing from the present
XX invention, which correspond to SEQ ID NO:1 to 2815, and then the last
XX 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
XX differ from the previously named sequences. SEQ ID NO:11 to 1680

CC (AAA23233 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.

XX
SQ Sequence 3490 BP; 880 A; 966 C; 872 G; 772 T; 0 other;

Alignment Scores:
Pred. No.: 7.68e-115 Length: 3490
Score: 1574.00 Matches: 313
Percent Similarity: 99.37% Conservative: 1
Best Local Similarity: 99.05% Mismatches: 2
Query Match: 98.93% Indels: 0
DB: 21 Gaps: 0

US-09-836-602-4 (1-316) x AAA34777 (1-3490)

Qy 1 MetProArgGlyTrpPhrAlaLeuCysLeuLeuSerLeuProSerGlyPheMetSer 20
Db 1169 ATCCCGGGGGCTGGACCGCGCTTCTGCTGAGTTGCTGCTTCTGGGTTTCATGAGT 1228
Qy 21 LeuAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 1229 CTTGACAAACAGCGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 1288
Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 1289 TCTACAAATGTATCTTACCAAGAACTACAAACCTAGTACCTTGGAGTACCAAGCTG 1348
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
Db 1349 CAGCCTGTGTCTCAACATGGCAATGAGGCCACAAACATCACAGAAAGACAGTCAA 1408
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 1409 TTTACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAACTCTTCTGTCCAGTCA 1468
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 1469 CAGACCTCTGTATCAGCAGAGTTTACCACCCAGCCCAACCTTCACTCCAGAGACA 1528
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 1529 ACCTTGAAGCTAGCCTGTCCACCTGGAAATGTTTCAGACCTTTCACACCTAGCATTAGC 1588
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 1589 CTTGCAACATCTCCCATTAACCTATACATCATCTTCTCTATCCTAAGTGACATCAAG 1648
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 1649 GCAGAAATCAAAATGTTCCAGGCATCAGAGAGTCAAAATTGACTCAGGCGCATCTGCCTGGAG 1708
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 1709 CAAATAAAGACCTCCAGCTGTGGGAGTTTAAAGAGACAGAGGGGAGGCGCTGGCCGA 1768
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 1769 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGATGCTGCTGGGCGCCAGGTATGCTCCCGTCTC 1828
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu 240
Db 1829 CTTGCCAGTCTGAGGTGAGGCGCTCAGTGTCTACTGCTGTGCTTGGCCACACAGACAGAA 1888
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 1889 ATTTCCAGCAAACTCCAACTTATGAAAGACCACTCTGACCTGAAAGAGCTGGGATC 1948
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuLe 280
Db 1949 CTAGATTTTACTGAGCAAGATGTTGCAAGCCACAGAGCTATTTCCTCCAAAGAGCCCTGATT 2008

Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 2009 GCACGTGTCACCTCGGAGCCCTGCTGGCTGCTGCTGGCATCATCTTCTCTGATG 2068
Qy 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro 316
Db 2069 AATCGCCGAGCTGGAGCCGCCACAGAGAAAGGCTGGCGGAGACCTT 2116
RESULT 8
AAD22663
ID AAD22663 standard; DNA; 906 BP.
XX AAD22663;
AC AAD22663;
XX
DT 26-FEB-2002 (first entry)
XX Human CD34 surface antigen deleted variant (dCD34) encoding DNA.
DE Human; surface marker; surface antigen; T lymphocyte; gene therapy;
XX CD34 deleted variant; dCD34; ds.
KW Homo sapiens.
OS Synthetic.
FH Key
CDS 1..906
FT /*tag= a
FT /product= "Human CD34 surface antigen
FT deleted variant (dCD34)"
XX EP1148066-A1.
PN 24-OCT-2001.
XX
PD 18-APR-2001; 2001EP-0109374.
PF 18-APR-2000; 2000DE-1019075.
PR (ZAND/) ZANDER A R.
PA Zander AR;
PI
XX WPI; 2002-019289/03.
DR P-PSDB; AAE13543.
XX New gene transfer vector (accession number DSM13396) containing a
transgene and a nucleic acid sequence coding for a surface marker,
useful in gene therapy, and for detecting genetically modified cells or
cells which do not express CD34 -
PS Claim 3; Page 14-15; 28pp; English.
XX The patent discloses a gene transfer vector (accession number DSM13396)
containing a transgene and a nucleic acid sequence coding for a surface
marker. The surface marker is the CD34 surface antigen, its fragment or
variant. The vector is useful for in vitro transduction of T lymphocytes,
for gene therapy, and in the enrichment, detection and analysis of cells
in vitro that do not naturally express CD34. T lymphocytes transduced
with the vector are also useful in gene therapy. The CD34 nucleic acid
sequences (marker genes), their fragments or variants are used for
detecting genetically modified cells or cells which do not naturally
express CD34. The present sequence is a DNA encoding human CD34 surface
antigen deleted variant (dCD34).
SQ Sequence 906 BP; 247 A; 250 C; 199 G; 210 T; 0 other;
Alignment Scores:
Pred. No.: 1.84e-110 Length: 906
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 94.85% Indels: 0
DB: 24 Gaps: 0

US-09-836-602-4 (1-316) x AAD22663 (1-906)

```
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 1 ATGCCGCGGGGTGGACCGCGCTTGTGCTGTGAGTTGTGCTGCCCTCTGGGTTTCATGAGT 60
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 61 CTTGACAACAACGGTACTGCTACCCCGACAGTTACTACCCAGGGAACATTTTCAAAATGTT 120
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 121 TCTCAAAATGATCTTACCAAGAACTACACACCTAGTACCTTGGAGTACACGCGCTG 180
Qy 61 HisProValSerGlnHisGlyAsnGluLaThrThrAsnIleThrGluThrThrValLys 80
Db 181 CACCCTGTGTCTCAACATGGCAATGAGGCCACACAAACATCACAGAAACGACAGTCAAA 240
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 241 TTCACATCTACCTCTGTGATACCTTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 300
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 301 CAGACCTCTGTAATCAGCACAGTGTTCACCAACCCAGCCAGCTTCAACTCCACAGACA 360
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 361 ACCTTGAAGCCCTAGCCTGCTACCTGGAAATGTTTCAGACCTTTCACCACTAGCCTAGC 420
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 421 CTTGCAACATCTCCCACTAAACCCATACATCATCTTCTCCATCTTAAGTGACATCAAG 480
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 481 GCAGAAATCAAAATGTTCAAGGCATCAGAGAAGTGAATTGACTCAGGGCATCTGCCTGGAG 540
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlyLeuAlaArg 200
Db 541 CAAAATAAGACCTCCAGCTGTGCGAGTGTAAAGAGGACAGGGGAGGGGCTGGCCCGA 600
Qy 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 601 GTGCTGTGTGGGAGGAGCAGCTGATGCTGATGCTGGGGCCAGGTATGCTCCCTGCTC 660
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu 240
Db 661 CTTGCCCATCTGAGTGAGGCTCAGTGTCTACTGCTGTGCTTGGCCCAACAGACAGAA 720
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 721 ATTTCCAGAACTCCAATTTATGAAAGACCAATCTGACCTGAAAAGCTGGGGATC 780
Qy 261 LeuAspPheThrGlnGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 781 CTAGATTTCACCTGACAGAGATGTTCCAGCCACAGAGCTATTTCCAAAAGACCTGATT 840
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 841 GCACCTGGTCACCTCGGGAGCCCTGCTGGCTGTCTTGGGCATCACTGGCTATTTCCTGATG 900
Qy 301 Asn 301
Db 901 AAT 903
RESULT 9
AAT59508
ID AAT59508 standard; cDNA; 1065 BP.
XX
AC
XX AAT59508;
XX
```

```
DT 27-OCT-1997 (first entry)
XX
DE Porcine CD34 from clone 1AX2-3 encoding cDNA.
XX
KW Antibody; bone marrow chimera; graft; antigen; pig; immune response;
KW haematopoietic progenitor cell; stem cell; cord blood; ss.
XX
OS Sus scrofa.
XX
FH Key Location/Qualifiers
CDS 31..1065
FT /*tag= a
FT /product= CD34
FT sig_peptide 31..123
FT /*tag= b
FT mat_peptide 124..1062
FT /*tag= c
XX
PN WO9640244-A1.
XX
PD 19-DEC-1996.
XX
PF 03-JUN-1996; 96WO-US08340.
XX
PR 07-JUN-1995; 95US-0475634.
XX
PA (BIOT-) BIOTRANSPLANT INC.
XX
PI Hawley RJ, Monroy RL;
XX P-PSDB; AAW11822.
XX
DR WPI; 1997-108628/10.
XX
PT Antibodies selective for CD34 + porcine cells - used to isolate such
PT cells for generating bone marrow chimera(s) that will accept porcine
PT grafts, also new porcine CD34 antigen
XX
PS Example 1; Fig 1; 35pp; English.
XX
CC The present sequence encodes porcine CD34. Antibodies (Ab) have been
CC produced that recognise porcine CD34+ cells. The Ab's are used to
CC recover CD34+ cells, preferably from porcine bone marrow but also from
CC cord blood, and CD34+ cells, which are enriched in haematopoietic
CC progenitor cells, especially stem cells, are used to generate bone
CC marrow chimeras in a human to assist acceptance of a porcine graft
CC (i.e. to prevent or inhibit an immune response against the graft).
CC Ab's can also be used to detect CD34+ cells, e.g. to determine if
CC mixed chimerism has been established. The CD34 polypeptide and its
CC fragments are used to raise Ab.
XX
SQ Sequence 1065 BP; 248 A; 325 C; 273 G; 219 T; 0 other;
Alignment Scores:
Pred. NO.: 7.13e-64 Length: 1065
Score: 921.00 Matches: 199
Percent Similarity: 70.03% Conservative: 37
Best Local Similarity: 59.05% Mismatches: 75
Query Match: 57.89% Indels: 26
DB: 18 Gaps: 6
US-09-836-602-4 (1-316) x AAT59508 (1-1065)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 67 ATGCCGCGGGGTGGACCGCTGCTGCTGTGAGTTGTGCTGCCCTCTGGGTTTCACAGT 126
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuPro----- 32
Db 127 GTG---AACAGCTCAACTATTGCTTCCACCTTGCACGCTGCCGCTGGGTCAACTCCACC 183
Qy 33 -----ThrGlnGlyThrPheSerAsn----- 39
Db 184 GGGCGCGCTACCGAGGGCGACGTATCACCAGGTCACACTATCTCAGACATATCTTCACT 243
```


ID ABA51134 standard; DNA; 233 BP.
XX
AC ABA51134;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #9829.
DE
DE
KW Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 4; SEQ ID NO 9829; 327pp + sequence listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and Bt 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at fcp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
Pred. No.: 1,71e-16 Length: 233
Score: 311.00 Matches: 60
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.55% Indels: 0
DB: 22 Gaps: 0

US-09-836-602-4 (1-316) x ABA51134 (1-233)

QY 16 SerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGly 35
DB 231 TCtGGGTTCATGAGTCTTGACACACACGGTACTGCTACCCAGAGTTACCTACCCAGGGA 172

QY 36 ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeu 55
DB 171 ACATTTTCAAAATGTTCTACAAATGTATCTACCAAGAACTACAAACACCTAGTACCCCT 112
QY 56 GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr 75
DB 111 GGAAGTACCAGCTGCACCTGTGTCTCAACATGGCAATGAGGCCACACAAACATCACA 52
RESULT 12
ABA69132/c
ID ABA69132 standard; DNA; 233 BP.
XX
XX ABA69132;
AC
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #17437.
DE
DE
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
KW
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
XX 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX Claim 4; SEQ ID NO 17437; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at fcp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
Pred. No.: 1,71e-16 Length: 233
Score: 311.00 Matches: 60
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.55% Indels: 0
DB: 22 Gaps: 0

US-09-836-602-4 (1-316) x ABA69132 (1-233)

QY 16 SerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGly 35
DB 231 TCtGGGTTCATGAGTCTTGACACACACGGTACTGCTACCCAGAGTTACCTACCCAGGGA 172

US-09-836-602-4 (1-316) x AAK17442 (1-233)

Qy	16	SerGlyPheMetSerLeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGly	35
Db	231	TCTGGGTTCATGAGTCTTGACAAACGGTACTGCTACCCAGAGTTACCTACCCAGGGA	172
Qy	36	ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeu	55
Db	171	ACAATTTTCAAAATGTTTCTACAAATGTATCCTACCAAGAAACTACAACACTAGTACCCCTT	112
Qy	56	GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr	75
Db	111	GGAAGTACACAGCTGCACCTGTGCTCTCAACATGGCAATGAGGCCACAACAACATCACA	52

RESULT 15

AAK43242/C

ID AAK43242 standard: DNA: 233 BP.

AA
AC AAK43242:XX
DT 06-NOV-20

XX

XX

KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

[illegible]XX
ново зартема:

PN WO20015/2/6-A2.
YY

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.

PR 30-JUN-2000; 2000DS-0608408.

PK 03-AUG-2000; 2000US-0632388;
PB 21-SEP-2000; 2000US-0234687;

PR 27-SEP-2000; 2000S-0236359.
PR 04-OCT-2000; 2000CB-0031263

XX

XX

XX
XX

XX
XX

PT analyzing gene expression in human bone marrow genome-derived single exon nucleotides

XX
DC
Example 1: CEO TD NO[illegible]

CC probes which are derived from genomic sequences expressed in the human

CC samples, which may enable the improved diagnosis and treatment

CC such as lymphoma) leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.

Sequence 233 BP: 55 A: 47 C: 61 G: 70 T: 0 other:

Alignment Scores:

Pred. No.:	1.71e-16	Length:	233
Count:	217	Count:	20

Percent Similarity: 100.00% Conservative: 0

Query Match:	19.55%	Indels:	0
--------------	--------	---------	---

52

(CCZ-T) 747C4VH Y (OTC-T) 4-200-058.60.60

[illegible]

Search completed: October 30, 2002. 08:25:46

Job time : 152.956 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:08:08 ; Search time 1619.9 Seconds
(without alignments)
4082.221 Million cell updates/sec

Title: US-09-836-602-4
Perfect score: 1591
Sequence: 1 MPRGWTALCLLSLLPSGFMS.....YFLMNRKSWSTGRLLELP 316

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_lj/USPTO_spool/US09836602/runat_29102002_101144_3785/app_query.fasta_1.1429
-DB=GenEmbl -QMT=fastcap -SUFFIX=rgc -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602.ecgn_1.1.3800 -runat_29102002_101144_3785 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_XMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*

- 29: em_vl.*
- 30: em_hgt_hum.*
- 31: em_hgt_inv.*
- 32: em_hgt_other.*
- 33: em_hgt_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1591	100.0	2657	6	AX333753	AX333753 Sequence
2	1591	100.0	2657	9	S53911	S53911 CD34-glycop
3	1574	98.9	2463	9	S53910	S53910 CD34-glycop
4	1574	98.9	2615	9	HUMCD34HS	M81104 Human CD34
5	964	60.6	2956	4	CFU49457	U49457 Canis faml
6	921	57.9	1065	6	AR078454	AR078454 Sequence
7	915	57.5	2690	4	AF461503	AF461503 Sus scrofa
8	887.5	55.8	978	10	S69301	S69301 CD34-cell s
9	870.5	54.7	1149	10	S69299	S69299 CD34-cell s
10	870.5	54.7	1260	10	S69293	S69293 CD34-stem c
11	870.5	54.7	2398	10	BC006607	BC006607 Mus muscu
12	793	49.8	581	9	AF202879	AF202879 Homo sapi
13	772.5	48.6	2453	4	AB021662	AB021662 Bos tauru
14	741.5	46.6	2420	4	AF259378	AF259378 Capra hlr
15	617	38.8	160771	9	HS88L2	AL035091 Human DNA
16	617	38.8	182742	2	AC084393	AC084393 Homo sapi
17	617	38.8	185162	2	AL365178	AL365178 Homo sapi
18	617	38.8	212535	2	AL356275	AL356275 Homo sapi
19	409	25.7	254	9	HUMCD34S3	M81940 Human CD34
20	374.5	23.5	98347	2	AL513203	AL513203 Mus muscu
21	374.5	23.5	166333	2	AL513470	AL513470 Mus muscu
22	311	19.5	183	9	HUMCD34S2	M81939 Human CD34
23	275	17.3	165	9	HUMCD34S7	M81942 Human CD34
24	260	16.3	157	9	HUMCD34S5	M81944 Human CD34
25	233	14.6	235	6	AX247364	AX247364 Sequence
26	172.5	10.8	3124	5	MMTHRAD1	Y13978 Gallus gall
27	167	10.5	80272	5	AL591180	AL591180 Zebrafish
28	166	10.4	89232	5	AL591175	AL591175 Zebrafish
29	162.5	10.2	55069	8	Y5CH9196	U11583 Saccharomyc
30	162.5	10.2	184427	14	EHVU20824	U20824 Equine herp
31	160.5	10.1	253273	2	AL669823	AL669823 Mus muscu
32	158.5	10.0	183358	9	AC019041	AC019041 Homo sapi
33	156.5	9.8	2161	5	XELFIMC1X	L02115 Frog intege
34	155.5	9.8	167390	9	AC007263	AC007263 Homo sapi
35	155.5	9.8	182534	10	AC074046	AC074046 Mus muscu
36	154	9.7	3053	8	SCYNR044W	Z71659 S.cerevisia
37	154	9.7	3137	8	YSCAAGLCS	M60590 S.cerevisia
38	153.5	9.6	167254	9	CNS059DS	AL357093 Human chr
39	153.5	9.6	172789	2	AC104020	AC104020 Homo sapi
40	153	9.6	165227	2	AC098014	AC098014 Rattus no
41	153	9.6	185337	2	AC095762	AC095762 Rattus no
42	153	9.6	286060	2	AL589862	AL589862 Homo sapi
43	152.5	9.6	196553	2	AC073946	AC073946 Mus muscu
44	152	9.6	12412	3	CELH43E16	AC006669 Caenorhab
45	152	9.6	131274	2	AC096869	AC096869 Rattus no

ALIGNMENTS

RESULT 1
AX333753
LOCUS AX333753 2657 bp DNA
DEFINITION Sequence 4262 from Patent WO0194629.
ACCESSION AX333753
VERSION AX333753.1 GI:18124472
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (sites)

REFERENCE
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G., Horrigan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature gene sets
JOURNAL Patent: WO 0194629-A 4262 13-DEC-2001;
Avalon Pharmaceuticals (US)

FEATURES
source Location/Qualifiers
1. .2657
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 624 a 790 c 609 g 634 t
ORIGIN

Alignment Scores:
Pred. No.: 2.58e-97 Length: 2657
Score: 1591.00 Matches: 316
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-836-602-4 (1-316) x AX333753 (1-2657)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
DB 127 ATGCCGGGGGCTGGACCGGCTTTCGCTGAGTTGCTGCTTGGGTTTCATAGT 186

QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
DB 187 CTTGACAAACAGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTCAAATGTT 246

QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
DB 247 TCACAAATATCTCTACCAAGAACTACACACCTAGTACCTTGGAGTACGAGCTG 306

QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
DB 307 CACCCTGTGCTCAACATGGCATGAGGCCACAAACATCACAGAAACGACAGTCAA 366

QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
DB 367 TTCACATCTACCTGTGTGATAAGCTCAGTTATGGAAACACAAACTTCTGTCTCCAGTCA 426

QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
DB 427 CAGACCTCTGTATCAGCAGAGTGTCCACCACCCAGCAACGTTTCAACTCCAGAGACA 486

QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
DB 487 ACCTTGAAGCCTAGCCCTGTACACCTGGAAATGTTTCAGACCTTTCACACCTAGCCTAGC 546

QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
DB 547 CTTGCAACATCTCCCACTAAACCCCTATACATCATCTCTCTCTATCTCTAAGTGCATCAAG 606

QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
DB 607 GCAGAAATCAATGTTTCAGGCATCAGAGAGTGAATTTGACTCAGGGCATCTGCCCTGGAG 666

QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
DB 667 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAGACAGGAGGAGGCGCTGGCCCGA 726

QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
DB 727 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGCTGGGCCCGCAGGTATGCTCCCTGCTC 786

QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
DB 787 CTTGCCCGAGTCTCAGGTGAGGCGCTCAGTGTCTACTGCTGTGCTTTGGCCCAACAGACAG 846

QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
DB 847 ATTTCAGCAAACTCCCAACTTATGAAAAACACCACTCTGACCTGAAAAAGCTGGGATC 906

QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
DB 907 CTAGATTTCATCAGCAAGATGTTGCAAGCCACAGAGCTATTCCCAAAAGACCTGATT 966

QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
DB 967 GCATGTGTCACCTCGGAGCCCTGCTGCTCTTGGCATCATCTGGCTATTCTCTCATG 1026

QY 301 AsnArgArgSerTyrSerProThrGlyGluArgLeuGluLeuGluPro 316
DB 1027 AATCGCCGACGCTGGAGCCCAACAGGAAAGGCTGGAGCTGGACCC 1074

RESULT 2
S53911
LOCUS
DEFINITION
CD34=glycoprotein expressed in lymphohematopoietic progenitor cells
(alternatively spliced, truncated form) [human, UT7, mRNA, 2657
nt].
ACCESSION S53911
VERSION S53911.1 GI:264768
KEYWORDS human UT7.
SOURCE Homo sapiens
ORGANISM
REFERENCE
AUTHORS Nakamura,Y., Komano,H. and Nakauchi,H.
TITLE Two alternative forms of cDNA encoding CD34
JOURNAL Exp. Hematol. 21 (2), 236-242 (1993)
MEDLINE 93146100
REMARK Genbank staff at the National Library of Medicine created this
entry [NCBI gibbsq 124009] from the original journal article.
This sequence comes from Fig. 1AB.

FEATURES
Location/Qualifiers
1. .2657
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .2657
/gene="CD34"
91. .1077
/gene="CD34"
/note="glycoprotein expressed in lymphohematopoietic
progenitor cells; This sequence comes from Fig. 1AB"
/codon_start=1
/product="CD34"
/protein_id="AAB25223.1"
/db_xref="GI:264769"
/translation="MLVRRGARAGPMPRGWTALCLLSLPSEMSLDNNGTATPELP
TQGTNSVSTNYSYQETTPSLGSLHPVSOHQNEATNITETVKTSTSVTSV
YNTNSVSQSVISVETFTPANVSTPETTLKPSLSPGNVSDLSVTSLSLATSPKP
YTSSEPLSDIKAEIKCGIRVKTQGLKQNKTSQCAEFKDKRGEGLARVLCGEE
QADAGAQVCSLLAQSEVPQCLLVLANRTEISSKLQMKKHQSDKLKGLDFT
EQDVASHQSYSQKTLIALVTSGALLAVLGITCYFLNRRRSWSPTGERLEUP"

BASE COUNT 624 a 790 c 609 g 634 t
ORIGIN

Alignment Scores:
Pred. No.: 2.58e-97 Length: 2657
Score: 1591.00 Matches: 316
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-4 (1-316) x S53911 (1-2657)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
|||||


```
Db 127 ATGCCGCGGGCTGACCGCGCTTGGCTTGGCTGAGTTTGGCTTCTGGGTTCATGAGT 186
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 187 CTTGACACACAGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 246
Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 247 TCTACAAATGATCTACCAAGAACTACAACACCTAGTACCTTGGAGTACCAGCCTG 306
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 307 CACCCTGTGCTCAACATGGCAATGAGCCACAACAACATCACAGAACGACAGTCAAA 366
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAspSerSerValGlnSer 100
Db 367 TTCACATCTACCTCTGTGATACCTCAGTTATGGAACACAACTCTTCTGCCAGTCA 426
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 427 CAGACCTCTGTAAATCAGCACAGTGTACCACCCAGCCCAAGTTCACCTCACTACG 486
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThr 140
Db 487 ACCTTGAAGCCCTAGCCTGTCACTGGAAATGTTTCAGACCTTTCACCACTAGCCTAG 546
Qy 141 LeuAlaThrSerProThrLysProThrThrSerSerSerProIleLeuSerAspIleLys 160
Db 547 CTTGCAACATCTCCACCTAAACCTTATACATCTTCTCTATCTTAAGTACATCAAG 606
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 607 GCAGAAATCAATGTTTCAGGCATCAGAGAGTGAATTCACCTCAGGCGATCTGCCTGG 666
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 667 CAAAATAAGACCTCCAGCTGTGCGGAGTTTAAGAAGGACAGGAGGGCGCTGGCCCGA 726
Qy 201 ValLeuCysGlyGluGluAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 727 GTGCTGTGTGGGGAGGACGCTGATGCTGATGCTGGGGCCCGGATGCTGCTGCTGCTC 786
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu 240
Db 787 CTTGCCAGCTCAGGTGAGGCTCAGTGTCTACTGCTGCTGCTGCTGCCCAACAGACAG 846
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLeuGlyIle 260
Db 847 ATTTCCAGCAAACTCCAACTTATGAAAAGCACCACCAATCTGACCTGAAAAGCTGGGATC 906
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 907 CTGAGATTTCACGTGACAGATGTTGCAGCCACAGAGTATTCACCAAGAACCTTGATT 966
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 967 GCACCTGGTCACCTCGGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1026
Qy 301 AsnArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro 316
Db 1027 AATCGCCGAGCTGGAGCCCAAGGAAAGGCTGGAGCTGGAACCC 1074

RESULT 3
S53910
LOCUS
DEFINITION
S53910
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REMARK
FEATURES
source
gene
CDS
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:
US-09-836-602-4 (1-316) x S53910 (1-2463)
Qy 1 MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 127 ATGCCGCGGGCTGACCGCGCTTGGCTTGGCTGAGTTTGGCTTCTGGGTTCATGAGT 186
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 187 CTTGACACACAGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 246
Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 247 TCTACAAATGATCTACCAAGAACTACAACACCTAGTACCTTGGAGTACCAGCCTG 306
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 307 CACCCTGTGCTCAACATGGCAATGAGCCACAACAACATCACAGAACGACAGTCAAA 366
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 367 TTCACATCTACCTCTGTGATACCTCAGTTATGGAACACAACTCTTCTGCCAGTCA 426
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 427 CAGACCTCTGTAAATCAGCACAGTGTACCACCCAGCCCAAGTTCACCTCACTACG 486
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThr 140
Db 487 ACCTTGAAGCCCTAGCCTGTCACTGGAAATGTTTCAGACCTTTCACCACTAGCCTAG 546
Qy 141 LeuAlaThrSerProThrLysProThrThrSerSerSerProIleLeuSerAspIleLys 160
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2463)
Nakamura, Y., Komano, H. and Nakachi, H.
Two alternative forms of cDNA encoding CD34
Exp. Hematol. 21 (2), 236-242 (1993)
93146100
GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 124007] from the original journal article.
This sequence comes from fig. 1A.
Location/Qualifiers
1..2463
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..2463
/gene="CD34"
91..1248
/gene="CD34"
/note="glycoprotein expressed in lymphohematopoietic
progenitor cells; This sequence comes from fig. 1A"
/codon_start=1
/product="CD34"
/protein_id="AAB25222.1"
/db_xref="GI:264767"
translation="MLVRRGARAGPRMGRWTALCLLLSPGMSLDNNGTATPELP
TQGTFSNVSTNVSYQETTPSTLSTLHPSVQHNGEATNTTETVATSTVTSVTS
YGTNTSSVQSOTSVISTVTPANVSTPETTLKPSLSPGVSDLTSTSTSLATPTKP
YTSPPILSDIKAEIKCSGIREVKLTGICLQKNTSSCAFEKDRGELARVLGEE
QADADAGNQVCSLLLAQSEVRPQCLLVLANRTEISKKQLMKHQSDKLKGLDFT
EDGVASHQSYSGTSLIALVTSGLALLVIGITGYFLMNRSSWSQGERGEDPYTENG
GGGYSGGPSTPEAQKASVNRQAQENGTCQATSRNGHSARQHVVADEL"
```


LOCUS CFU49457 2956 bp mRNA linear MAM 21-JAN-1997
DEFINITION Canis familiaris hematopoietic progenitor cell marker CD34 mRNA,
complete cds.
ACCESSION U49457
VERSION U49457.1 GI:1224105
KEYWORDS
SOURCE
ORGANISM
dog.
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 2956)
AUTHORS McSweeney, P.A., Rouleau, K.A., Storb, R., Bolles, L., Wallace, P.M.,
Beauchamp, M., Krizanac-Bengez, L., Moore, P., Sale, G., Sandmaier, B.,
de Revel, T., Appelbaum, F.R. and Nash, R.A.
TITLE Canine CD34: cloning of the cDNA and evaluation of an antiserum to
recombinant protein
JOURNAL Blood 88 (6), 1992-2003 (1996)
MEDLINE 96420219
REFERENCE 2 (bases 1 to 2956)
AUTHORS McSweeney, P.A.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-1996) P.A. McSweeney, Transplant Biology, FHCR,
1124 Columbia St, Seattle, WA 98104, USA
FEATURES
source
1..2956
/organism="Canis familiaris"
/db_xref="taxon:9615"
/tissue_type="ML2 myelomonocytic leukemia"
1..2956
/gene="CD34"
260..1429
/gene="CD34"
/note="type 1 transmembrane protein"
/codon_start=1
/product="hematopoietic progenitor cell marker CD34"
/protein_id="AAB41055.1"
/db_xref="GI:1224106"
translation="MLAGRGARAGGLPRGWALCLLSLPFGFTNTVTPVT
STELSAVENTSKRAITLTPSGTTLVSQDSGGTATISFTVHVTSTSEILTL
PGTMSSVQSRLAITVFTPTNFSTSVTLPSLLPONGSDPPYNSLTSTPEY
YLSLPTPAPAGVSTLKEIKCSGVKVLNQICLELNSTSSCDPKKDKNEELTKQ
VICEKEPBRAGVSLLAQSEVPHCLLVLANKELEFSLQLLRKHQSLKILGI
RFTEDQVDSQSYSRKTLIALVLTGTYFLMNRNRSWSPGTGERLGEDPY
TENGGGQGYSSSGPESQAQGRKAVNRGPQENGTOATSRNGHSARQHVMADTEL"
BASE COUNT 625 a 864 c 815 g 652 t
ORIGIN

Alignment Scores:
Pred. No.: 2,17e-55 Length: 2956
Score: 964.00 Matches: 208
Percent Similarity: 72.98% Conservative: 27
Best Local Similarity: 64.60% Mismatches: 79
Query Match: 60.59% Indels: 2
DB: 4 Gaps: 2

US-09-836-602-4 (1-316) x CFU49457 (1-2956)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 296 CTGCGCGGGCTGGACCGCTCTGCTGCTCAGTCTGCTGCTTGGGTTTCACAAAC 355

QY 21 LeuAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 356 ACAGAAACCGTGATTACTCTACACAGTCCCAACCTCCACAGAAATAATGTCAGCTGTT 415

QY 41 SerThrAsnValSerTyGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 416 TCTGAGATACATCCAAACGGGAGCCATCACACTAACTCCTCTGGAACCTACCACCGTG 475

QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 476 TACTCTGTCTCAAGACAGCAGTGGGACCAACACCAACCATCTCAGAGACTACAGTCCAT 535

QY 81 PheThrSerThrSerValIleThrSerValTyThrSerValTyGlyAsnThrAsnSerSerValGlnSer 100
Db 536 GTCACATCTACCTCTGAGATCACCTTAACCGCTGGGACCATGAACCTCTCTGTTTCAGTCG 595

QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 596 CAGACCTCTTTAGCTATCAGCGTATCTTTTACCCCAACCACTTTTCAACTTCAAGTGTG 655

QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 656 ACCTTGGAGCCCGAGCTGCTACTCTGGAATGGTTCGATCCCTCCCTCAACAGCACCAGC 715

QY 141 LeuAlaThrSerProThrLysProTyThrSerSerPro----- 154
Db 716 CTGTGTGACATCCCCAGCGAATATTATACATCACTTCTCTACCCCAAGTAGAATGAC 775

QY 155 IleLeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThr 174
Db 776 ACCCAAGTACCATCAAGGGAGAAATCAAAATGTTCCGAGTCAAGAAGTGAATTTGAAC 835

QY 175 GlnGlyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArg 194
Db 836 CAAGGTATCTGCCTAGAGCTAAATGAGACCTCCAGCTGTGAGGACTTTAAGAAGATAAC 895

QY 195 GlyGluGlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAla 214
Db 896 GAGAAAACTGACCCCAAGTCTCTGTGT-----GAGAAGGAGCCAGCTGAGGCTGGGCC 949

QY 215 GlnValCysSerLeuLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuVal 234
Db 950 GGGGTGTGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1009

QY 235 LeuAlaAsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAsp 254
Db 1010 TTGGCCAAACAAAACAGAACTTTTCAGTAACTCCAACCTCTGAGAAAGCACCAGCTCTGAC 1069

QY 255 LeuLysLysLeuGlyIleLeuAspPheThrGluGlnAspValAlaLaserHisGlnSerTyr 274
Db 1070 CTGAAAAGCTGGGATCCGAGACTTCACTGAACAAGATGTTGGGAGCCACAGAGCTAT 1129

QY 275 SerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIle 294
Db 1130 TCCCGAAGACCTGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1189

QY 295 ThrGlyTyrPheLeuMetAsnArgArgSerTrpSerProThrGlyGluArgLeuGluLeu 314
Db 1190 ACTGGTTACTTCTGTGATGAACCGCGCGAGTTGGAGCCCTACAGAGAAAGGCTGGGCGAA 1249

QY 315 GluPro 316
Db 1250 GACCCT 1255

RESULT 6
AR078454
LOCUS AR078454
DEFINITION Sequence 18 from patent US 5963644.
ACCESSION AR078454
VERSION AR078454.1 GI:10005200
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1065)
AUTHORS Hawley, R.J. and Monroy, R.L.
TITLE Porcine CD34
JOURNAL Patent: US 5963644-A 18 05-OCT-1999;
FEATURES Location/Qualifiers
source 1..1065
BASE COUNT 248 a 326 c 273 g 218 t
ORIGIN
Alignment Scores:

Pred. No.: 4.37e-53 Length: 1065
Score: 921.00 Matches: 199
Percent Similarity: 70.03% Conservative: 37
Best Local Similarity: 59.05% Mismatches: 75
Query Match: 57.89% Indels: 26
DB: 6 Gaps: 6

US-09-836-602-4 (1-316) x AR078454 (1-1065)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 67 ATGCCCGGGGGTGGACACAGCTGCTGCTGAGTGGCTCCCTCGGGTTCACAGCT 126
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuPro----- 32
Db 127 GTG---AACAGCTCAACTATTGCTTCCACCTTGCCAGCTGCCGCTGGGTCAACTCCAC 183
Qy 33 -----ThrGlnGlyThrPheSerAsn----- 39
Db 184 GGGCGCGGTACCGAGGGGACGCTATCACCGGGTCAACTATCTCAGACATATCTTCACCT 243
Qy 40 ValSerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSer 59
Db 244 GTTCTTACAAATATATCCACAGGAAACACATCA---GATGCTTTCGAAAGTCCAGC 300
Qy 60 LeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVal 79
Db 301 CTCACACTGTCTCAGGGCAGCAGTGGGACACCGTAGCATCTCAGGCCCTACAGTT 360
Qy 80 LysPheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGln 99
Db 361 AATTTCATGTCTACCTCGGGGTGCTACCTCGTCCCGAAACGGTTAACTCTTCTCTCCAG 420
Qy 100 SerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGlu 119
Db 421 CCTCAGACTCT--CTAGCCACACGCTCTCCGGCCACATCACTTTACAACTTCAGAG 477
Qy 120 ThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrThr 139
Db 478 GTGACCTGTCAGCCAGCAGCAGTTCAGGAAATGTTTTCAGACCCCTCTACACAGTACC 537
Qy 140 SerLeuAlaThrSerProThrLysProThrLysProThrSerSerProIleLeuSerAspIle 159
Db 538 AGCCCTGCGAGATCCCGCCACAGCCCTACACATCATCTCTCTCTACCCAGGTAGCCAC 597
Qy 160 LysAlaGluLeuLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeu 179
Db 598 AAGGGGGAAGTCAATGTGCCCCAAATCAAGAGAGGTGAATGACCCAGGTATCTGCCTG 657
Qy 180 GluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlyGluLeuAla 199
Db 658 GAGCGAAATGAGACCTCCGGCTGCGGAGATTTAAGAGGACAAATGGAGAGAGTTGATG 717
Qy 200 ArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeu 219
Db 718 CAAATCTCTGTGGCAGGAGCAGCTGAGCGCGGCCAGG-----GTGTGCTCCTTG 771
Qy 220 LeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThr 239
Db 772 CTCCTTGCCCAATCTGAGGTGAACCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 831
Qy 240 GluIleSerSerLysLeuGlnMetLysLysHisGlnSerAspIleLysLysLeuGly 259
Db 832 GAACCTAGCAGCAAGTCTCTGCTTCTGGAAGACACAGCTGTAAGTACGAGAGATGAGC 891
Qy 260 IleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeu 279
Db 892 ATCCAAAACTTCTCGAAACAAGATGTTAGAGGCCACACAGAGTACTCCGGAAGACCTTG 951
Qy 280 IleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeu 299
Db 952 ATTGCATCTGGTACCTCGGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1011

Qy 300 MetAsnArgArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro 316
Db 1012 ATGAACCGTTCGAGTTGGAGCCCTACAGGAGAAAGCTGGAGCTGGAACCC 1062

RESULT 7
AF461503
LOCUS AF461503 2690 bp mRNA linear MAM 24-JAN-2002
DEFINITION Sus scrofa CD34 antigen mRNA, complete cds.
ACCESSION AF461503
VERSION AF461503.1 GI:18308137
KEYWORDS
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 2690)
AUTHORS Sun, J. and Butler, J.E.
TITLE Cloning, sequencing and expression of swine CD34 gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2690)
AUTHORS Sun, J., Wang, J. and Butler, J.E.
TITLE Direct Submision
JOURNAL Submitted (20-DEC-2001) Microbiology, University of Iowa, 51 Newton
Road, Iowa City, IA 52242, USA
FEATURES
source
1..2690
/organism="Sus scrofa"
/db_xref="taxon:9823"
/tissue_type="mesenteric lymph nodes"
298..1467
/note="cluster of differentiation antigen 34"
/codon_start=1
/product="CD34 antigen"
/protein_id="AAL67838.1"
/db_xref="GI:18308138"
/translation="MPRGWTLCLLSLPFGTAVNSSTVASTLPTAAGSTPTGPATA
GTATGSAIDISPSVNTNISNEETPDAPESASLHTVSQSGSCTVAISGPTVNFMS
TSATLVPETVNNSVQPTSLATVSSATINFTSEVTLQPTFPFGNVSDPLNSTSPA
RSPTSPVTSPTPGSHKQCAQIKVKLTQIGICLERNETSGCEKFKDNGKLMQ
ILCGEOAEAGPGVCSLLLAQSEVKPHCLLVLANGTSLSKFLLEKHOSLEKREMSI
HFGSKQDVRSHQSVSRKTLIALVTSGILLAVLGTGYLLMNRSSWSTGBRGLGDPY
TENGSGGQYSSGPGASPEAGKASVTGKAQENGQTGATSRNGHSAROPMVADEL"

BASE COUNT 547 a 808 c 738 g 597 t
ORIGIN

Alignment Scores:
Pred. No.: 3.6e-52 Length: 2690
Score: 915.00 Matches: 199
Percent Similarity: 69.44% Conservative: 35
Best Local Similarity: 59.05% Mismatches: 77
Query Match: 57.51% Indels: 26
DB: 4 Gaps: 6

US-09-836-602-4 (1-316) x AF461503 (1-2690)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 298 ATGCCCGGGGGTGGACACAGCTGCTGCTGAGTGGCTCCCTCGGGTTCACAGCT 357
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuPro----- 32
Db 358 GTG---AACAGCTCAGCTGTTGCTTCCACCTTGCCCAACGCTGGGTCAACTCCAC 414
Qy 33 -----ThrGlnGlyThr-----PheSerAsn 39
Db 415 GGGCAGACTACGAGGACAGCATATCACCGGTACAGATATCTTCACCT 474
Qy 40 ValSerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSer 59
Db 475 GTTCTTACAAATATATCCACAGAGAA---ACCACACAGATGCTTTCGAAAGTCCAGC 531
Qy 60 LeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVal 79
Db 1011

```
Db 532 CTCACACTGTCTCTCAGGGCAGCAGTGGGACCACCGTAGCCATCTCAGGCCCTCAGATT 591
Qy 80 LysPheThrSerThrSerValIleThrSerValThrGlyAsnThrAsnSerSerValGln 99
Db 592 AATTTCATGTCTACCTCGCGCTACCTCGTCCCGGAAACGCTGAACCTTCTCTCCAG 651
Qy 100 SerGlnThrSerValIleSerValPheThrThrProAlaAsnValSerThrProGlu 119
Db 652 CCTGAGACTCT--CTAGCCACAGTGCCTCGGCCACCATCACTTTACAACTTCAGAG 708
Qy 120 ThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThr 139
Db 709 GTGACCTGCGACCCAGCAGCTTCCAGGAAATGTTTCCAGACCCCTCTACAACAGTACC 768
Qy 140 SerLeuAlaThrSerProThrLysProThrThrSerSerSerProIleLeuSerAspIle 159
Db 769 AGCCCTGCGAGATCCCCACCAAGCCCTTACACATCATCTCTCTCCACCCAGGTAGCCAC 828
Qy 160 LysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeu 179
Db 829 AAGGGGAGTCAATGTGCCAATCAAGAGGTGAATTTGACCCAGGTATCTGCCTG 888
Qy 180 GluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla 199
Db 889 GAGCGAAATGAGACCTCCGCTGCGAGAAGTTTAAAGAGGACAATGGAGAGAAGTTGATG 948
Qy 200 ArgValLeuCysGlyGluGlnAlaAspAlaGlyAlaGlnValCysSerLeu 219
Db 949 CAATTCCTGTGTGGCAGGAGCAGCTGAGCGCGGCCAGGG-----GTGTGCTCCTTG 1002
Qy 220 LeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThr 239
Db 1003 CTCCTTGCCCAATCTGAGTGAACCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1062
Qy 240 GluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGly 259
Db 1063 GAACTTAGCAGCAAGTCTGCTGCTGGAAGACCACTGCTGAAGTGAAGAGAGATGAGC 1122
Qy 260 IleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerThrSerGlnLysThrLeu 279
Db 1123 ATCCAACTCTCTGAAACAAGATGTTAGGACCAAGCAGACTCTCCGGAAGACCTTG 1182
Qy 280 IleAlaLeuValThrSerGlyAlaLeuAlaValLeuGlyIleThrGlyThrPheLeu 299
Db 1183 ATTGACCTGTCACCTCGGGGATCTGCTGGCTGCTGCTGGGATCAGTGGCTACTTGTCTG 1242
Qy 300 MetAsnArgArgSerTrpSerProThrGlyGluArgLeuGluLeuPro 316
Db 1243 ATGAACCGTCGAGTTGGAGCCCTACAGGAGAAAGCTGGCGAAGACCT 1293

RESULT 8
S69301
LOCUS S69301
DEFINITION CD34-cell surface antigen {alternatively spliced, clone 2} [mice,
stomal cell line PA-6, mRNA Partial, 978 nt].
ACCESSION S69301
VERSION S69301.1 GI:495715
KEYWORDS
SOURCE Mus sp. stomal cell line PA-6.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 978)
AUTHORS Suda,J., Sudo,T., Ito,M., Ohno,N., Yamaguchi,Y. and Suda,T.
TITLE Two types of murine CD34 mRNA generated by alternative splicing
JOURNAL Blood 79 (9), 2288-2295 (1992)
MEDLINE 92239883
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 145459] from the original Journal article.
This sequence comes from Fig. 3.
FEATURES
Location/Qualifiers
1..978
/organism="Mus sp."
```

```
gene
1..978
/partial
/contig="CD34"
CDS
1..978
/gene="CD34"
/feature="cell surface antigen; This sequence comes from Fig.
4"
/codon_start=1
/product="CD34"
/protein_id="AAB22109.1"
/db_xref="GI:495716"
/translation="MOVHRTDRAGLLLPWRVALCLMSLLHLLNLTSTSTQGIS
PSVTNSESVEENITSSIPGSTSHYLLIQDSKTTTPTAISETWNTVTSGIPSGSTPH
SSQPSPTGILPTTSDISTSEMTWKSLSPLSNVDSFNNKSEFEMTSPETPAYT
SSSAPAIKGIKSGIREVLAQICLSEASCEEFKEKEDLIIICEKEAE
ADAGASCVCLLAQSEVRPECLMLANSTELPSKQLMKRQSLDRKLIGTQSFNQD
IGSHQSYSRKTLIALVTSGVLLAILGTGTGYFLMNRSSWPTGERLELEP"
BASE COUNT 261 a 249 c 228 g 240 t
ORIGIN
Alignment Scores:
Pred. No.: 6,76e-51 Length: 978
Score: 887.50 Matches: 192
Percent Similarity: 72.50% Conservative: 40
Best Local Similarity: 60.00% Mismatches: 77
Query Match: 55.78% Indels: 11
DB: 10 Gaps: 4
US-09-836-602-4 (1-316) x S69301 (1-978)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 37 CTGCCATGGCGTGGTACCTCTGCTGCTGATGAGTCTGCTG-----CAT 81
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 82 CTAATAACTTGCATCTGCTACACGAGAGACTTCTACACAGGAATATCCCATCAGTT 141
Qy 41 SerThrAsnValSerTyrrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 142 CCTACCAATAGTCTGTTGAGGAAATATACATCTAGCATCCCTGGAAAGTACCAAGCC 201
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 202 TACTTGATCTATCAGACAGCAGTAAAGCACACAGCCATCTCAGAGACTATGGTCAAC 261
Qy 81 PheThrSerThrSerValIleThrSerValTyrrGlyAsnThrAsnSerSerValGlnSer 100
Db 262 TTTACAGTTACCTCTGGGATCCCTTCAGGCTCTGGAACTCCACACACTTTTTCACAACCA 321
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 322 CAGACTTCCCCAACTGGCATATGCTGCTACTTCTCAGACAGTATTTCCACTTCAGAGATG 381
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 382 ACCTGGAGTCCAGCTGCCATCTATAAATGTTCTGATTATTCGCTTAATAATAGCAGC 441
Qy 141 Leu---AlaThrSerProThrLysProTyrr-----ThrSerSerProIleLeu 156
Db 442 TTTGAGATCATATCACCACCGCCATATGCTTACACATCATCTCTCTCTCCG----- 495
Qy 157 SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
Db 496 AGTGCCATTAAAGGAGAAATCAATAGTCTCGAAATCCGAGAAGTGAAGTTGGCCAGGCT 555
Qy 177 IleCysLeuGlnAlaAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
Db 556 ATCTGCTGCTGAAGTAAAGTGAAGCATCTAGTTGTGAGGAGTTTAAAGAAAGGAGAGAA 615
Qy 197 GlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaGlyAlaGlnVal 216
```

```

Db      616  GATCTAATTCAAATACCTGTGTGAAGAGGAGGCTGAGGCTGATGCTGGTGTAGTGTC 675
Qy      217  CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
Db      676  TGCCTCCCTGCTTCTAGCCAGCTCTGAGTTAGGCTTGTGCTGTTGCTGATGGTCTTGCC 735
Qy      237  AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
Db      736  AATAGCAGACAACCTCCAGCAACTCCAGCTTATGAAAGACCAATCTGACTTGAGA 795
Qy      257  LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGln 276
Db      796  AAGCTGGGGATCCAAAGCTTCAATAAACAAGATATCGGAGCCACCAAGACTATTCCTCGA 855
Qy      277  LysThrLeuLeuAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGly 296
Db      856  AAGACTCTTATTGATTTGGTCACTCTGGAGTCTGCTGGCCATCTTGGCCACCACTGGT 915
Qy      297  TyrPheLeuMetAsnArgSerTyrSerProThrGlyGluArgLeuGluLeuPro 316
Db      916  TATTTCCTGATGAACCTCGCAGTTGGAGCCCTACAGGAGAAAGGCTGGAGCTGGAACCT 975

RESULT 9
S69299
LOCUS   CD34=cell surface antigen (alternatively spliced, clone 6-2-2)
DEFINITION
ACCESSION S69299
VERSION   S69299.1 GI:495713
KEYWORDS
SOURCE    Mus sp. stromal cell line PA-6.
ORGANISM  Mus sp.
REFERENCE
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
          1 (bases 1 to 1149)
TITLE     Suda,J., Sudo,T., Ito,M., Ohno,N., Yamaguchi,Y. and Suda,T.
          Two types of murine CD34 mRNA generated by alternative splicing
          JOURNAL Blood 79 (9), 2288-2295 (1992)
MEDLINE   92239883
REMARK    GenBank staff at the National Library of Medicine created this
          entry [NCBI gibbsq 145456] from the original journal article.
          This sequence comes from Fig. 3.
FEATURES
source    1..1149
           /organism="Mus sp."
           /db_xref="taxon:10095"
gene       1..1149
           /partial
           /gene="CD34"
CDS       1..1149
           /gene="CD34"
           /note="cell surface antigen; This sequence comes from Fig.
           3"
/codon_start=1
/product="CD34"
/protein_id="AAB22108.1"
/db_xref="GI:495714"
/translation="MOVHRDTRAGLLLPWRVALCMLSLHLNLTATTETSTOGIS
PSVPTNVEENITSSIPGTSYLIYQSSKTPAISVMVNFVTSGIPSGSGTGH
TSQPOTSPTGILPTTSDSISSEMFWKSLPSINVSIDSPNNSSEFMSPPEPYAT
SSAPAIKGEIKCSGIREVRLAQICLSEASSCEEFKKEKEDGLIQLICEKEAE
ADGASVCSLLQAQSVRPECLMLANSTELPSKLQLMKEHQSDRLKLGISFNKQD
IGHQSVKSTIALYTSGLIALLGTGYFLMNRRSWSPTRLERLGEDPYVTENGCGQ
GYSSGPGASPEQGNKAVTRGAQENCTGQATSRNHSARQHVVDTEL"
BASE COUNT 307 a 293 c 285 g 264 t
ORIGIN

Alignment Scores:
Pred. No.: 1,13e-49 Length: 1149
Score: 870.50 Matches: 189
Percent Similarity: 71.88% Conservative: 41
Best Local Similarity: 59.06% Mismatches: 79
Query Match: 54.71% Indels: 11

```

```

DB:      10      Gaps:      4
US-09-836-602-4 (1-316) x S69299 (1-1149)
Qy      1  MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db      37  CTGCCATGGCGCTGGGTAGCTCTCTGCTGATGAGTCTGCTG-----CAT 81
Qy      21  LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db      82  CTAATTAACCTTGACTTCTGCTACCGAGACTTCTACACAAGGAATATCCCCATCAGTT 141
Qy      41  SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db      142  CCTACCAATGAGTGTGTGAGGAAATATCACATCTAGCATCCCTGGAAGTACCAGCAC 201
Qy      61  HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db      202  TACTTGATCTATCAGGACAGCAGTAAGACCACACACCATCTCAGAGACTATGGTCAAC 261
Qy      81  PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db      262  TTTACAGTTACCTCTGGGATCCCTTCAGGCTCTGGAACCTCCACACACTTTTTCACACCA 321
Qy      101  GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db      322  CAGACTTCCCAACTGGCATCTGCTTACTTCTACACAGACTATTTCCTCCACTTCAGAGATG 381
Qy      121  ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db      382  ACCTGGAAGTCCAGCCTGCCATCTATAAATGTTTCTGATTATTTCGCTTAATAATAGCAG 441
Qy      141  Leu---AlaThrSerProThrLysProTyr-----ThrSerSerProIleLeu 156
Db      442  TTTGAGATGACATCACCACCGAGCCATATGCTTACACATCATCTCTGCTCCG----- 495
Qy      157  SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
Db      496  AGTGCCATTAAAGGAGAGAAATCAAAATGCTCTGGAATCCGAGAAGTGAGTGTGGCCAGGGT 555
Qy      177  IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
Db      556  ATCTGCCCTGGAACCTAAGTGAAGCATCTAGTTGTGAGAGATTAAAGAGAAAAGGGAGAA 615
Qy      197  GlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnVal 216
Db      616  GATCTAATTCAAATACCTGTGTGAAGAGGAGGCTGAGGCTGATGCTGGTGTAGTGTC 675
Qy      217  CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
Db      676  TGCCTCCCTGCTTCTAGCCAGCTGAGGTTAGGCCTGAGTGTGCTGCTGCTGGCC 735
Qy      237  AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
Db      736  AATAGCAGACAACCTTCCCAAGCAACTCCAGCTTATGAAAGACCAATCTGACTTGAGA 795
Qy      257  LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGln 276
Db      796  AAGCTGGGGATCCAAAGCTTCAATAAACAAGATATCGGAGCCACCAAGACTATTCCTCGA 855
Qy      277  LysThrLeuLeuAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGly 296
Db      856  AAGACTCTTATTGATTTGGTCACTCTGAGGATCTGCTGGCCATCTTGGCCACCACTGGT 915
Qy      297  TyrPheLeuMetAsnArgSerTyrSerProThrGlyGluArgLeuGluLeuPro 316
Db      916  TATTTCCTGATGAACCTCGCAGTTGGAGCCCTACAGGAGAAAGGCTGGGTGAAGACCT 975

RESULT 10
S69293
LOCUS   S69293
DEFINITION CD34=stem cell antigen [mice, bone marrow, mRNA, 1260 nt].
ACCESSION S69293

```

```

VERSION S69293.1 GI:495708
SOURCE Mus sp. bone marrow.
ORGANISM Mus sp.

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 1260)
Brown, J., Greaves, M.F. and Molgaard, H.V.
The gene encoding the stem cell antigen, CD34, is conserved in
mouse and expressed in haemopoietic progenitor cell lines, brain,
and embryonic fibroblasts
Int. Immunol. 3 (2), 175-184 (1991)
JOURNAL 91223042
MEDLINE GenBank staff at the National Library of Medicine created this
REMARK entry [NCBI gibbsq 145442] from the original journal article.
This sequence comes from Fig. 2.

FEATURES
source
1..1260
/organism="Mus sp."
/db_xref="taxon:10095"
gene
1..1260
/feature="CD34"
24..1172
/feature="CD34"
CDS
1..1260
/translation="MQVHRTFRAGLLPWRVWALCLMLLNLNLTATTTSTQGIS
PSVPTNFEVETITSSIPGSTSHYLIYQDSKTPPAISETMNFVTSGIPSGSGTGH
TFSPQSTPTGILPTTSDISITSEMTKSLPSINVSDYSPNNSFEMTSPTEPYAVT
SSAPSAIKGICSGIREVRLAOGICLSEASEEFKEKEDLIQILCEKEEAE
ADACSLKLLAQSVEPRCLMLVLANSTELPSKLOLMKEKHSDKLIGLQSPNKQD
IGSHQSVSRKTLIALYSGVLLAILCTTGTFMLNRRSWSPTEGLGDDPYTENGSGG
GYSGGSPPTQKANTVGAQENGTVQATSRNGHSAROHVADTEL"
BASE COUNT 325 a 321 c 320 g 294 t
ORIGIN

Alignment Scores:
Pred. No.: 1.29e-49 Length: 1260
Score: 870.50 Matches: 189
Percent Similarity: 71.88% Conservative: 41
Best local Similarity: 59.06% Mismatches: 79
Query Match: 54.71% Indels: 11
DB: 10 Gaps: 4

US-09-836-602-4 (1-316) x S69293 (1-1260)

Qy 1 MetProArgGlyThrAlaLeuLeuSerLeuLeuSerLeuLeuSerGlyPheMetSer 20
Db 60 CTGCCATGGCGTGGTACCTCTGCTGATGAGTCTGCTG-----CAT 104
Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 105 CTAAATAACTTACCTCTGCTACCGGAGACTTCTACACAGGAATATCCCATCAGTT 164
Qy 41 SerThrAsnValSerTyGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 165 OCTACCAATGAGTCTGTTCAGGAAATATACATCATCAGTCCCTCGGAGTACCGCCAC 224
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 225 TACTTGATCTATCAGACAGAGTAAAGACACACAGCCATCTCAGACTATGCTCAAC 284
Qy 81 PheThrSerThrSerValIleThrSerValTyGlyAsnThrAsnSerSerValGlnSer 100
Db 285 TTTACAGTTACCTCTGGGATCCCTCAGGCTCTGGAATCCACACACTTTTTCACACCA 344
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 345 CAGACTTCCCAACTGGGATAGTCTGCTACTTCTCAGACAGATATTTCCACTTCAGAGATG 404

```

121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
 141 Leu---AlaThrSerProThrLysProTyPyr-----ThrSerSerSerProIleLeu 156
 146 TTTGAGATGACATCACCCAGGAGCATATGCTTACACATCATCTTCTGCTCCG----- 518
 157 SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
 519 AGTGCCATTAAGGGAGAAATCAATGCTCTGGAATCCGAGAGTGGAGTGGCCAGGGT 578
 177 IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
 579 ATCTGCTGGAACTAAGTGAAGCATCTAGTTCTGAGGAGTTTAAAGAGGAAAGGAGAA 638
 197 GlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnVal 216
 639 GATCTAATTAATAACTGTGTGAAAGAGGAGGCTGAGGCTGATGCTGCTAGTCTGTC 698
 217 CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
 699 TGCTCCCTGCTCTAGCCAGTCTGAGGTAGGCTGAGTGTGCTGCTGATGCTGCTGCTG 758
 237 AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
 759 AATAGCACAGAACTTCCCGAGCAAACTCCAGCTTATGAAAGAGCACCACCAATCTGAC 818
 257 LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerThrSerGln 276
 819 AAGCTGGGAGTCCAAAGCTTCAATAAACAAGATATCGGAGCCAGCAGAGCTATTTCCGA 878
 277 LysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGly 296
 879 AAGACTTATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 938
 297 TyrPheLeuMetAsnArgSerTrpSerProThrGlyGluArgLeuGluLeuGluPro 316
 939 TATTTCTGATGACCGTCCGAGTTGGAGCCCTACAGGAGAAAGCTGGGTGAAGACCT 998

RESULT 11
 BC006607
 LOCUS Mus musculus, Similar to CD34 antigen, clone MGC:11720
 DEFINITION IMAGE:3966337, mRNA, complete cds.
 ACCESSION BC006607
 VERSION BC006607.1 GI:13879273
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 2398)
 Direct Submission
 Strausberg, R.
 Submitted (27-APR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: villalon@bcm.tmc.edu
 Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
 Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/INL at: <http://image.llnl.gov>
 Series: IRAK Plate: 17 Row: n Column: 8
 This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES

Location/Qualifiers

```

1..2398
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="MGC:11720 IMAGE:3966337"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCL CGAP_Mam1"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
62..1210
/codon_start=1
/product="Similar to CD34 antigen"
/protein_id="AAH06607.1"
/db_xref="GI:13879274"
/translation="MQVHRDRAGLLPWRVALCMLSLHLLNLTSTATERSTQGIS
PSVPTNESVEENITSPGSTSHYLYQDSKTPAISYETWNFTSGIPSGSTPH
TFSPQTSPTGILPTSDISTSEMTWKSLSPIVSDYSPNNSSEFMTSPTEPVAYT
SSAPASIKRGKCSGIRVRLAOGICELSPASSCEEFKEKGEDLIOLICEKEAE
ADAGASVCSILLAAQSVRPECLMLANSTELPSKIQLMKEKHQSLDKIKGIGSFNKOD
IGSHQSYSRKTLIALYTSVGLLAILGTTGYFLMNRWSWPTGERLGEDPYFYENGCGQ
GYSSGFGAFETQGRANVTGAGENGTGATSRNGHSARQHVVADTEL"

```

BASE COUNT

623 a 599 c 551 g 625 t

Alignment Scores:

```

Pred. No.:      2 91e-49      Length:      2398
Score:          870.50      Matches:      189
Percent Similarity: 71.88%      Conservative: 41
Best Local Similarity: 59.06%      Mismatches: 79
Query Match:      54.71%      Indels:      11
DB:              10          Gaps:         4

```

US-09-836-602-4 (1-316) x BC006607 (1-2398)

```

QY 1 MetProArgGlyThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98 CTGCCATGGCGCTGGTGGAGTCTCTGCTGATGAGTCTGCTG-----CAT 142

QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
   |||:||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 CTAATAAAGTGTGCTTCTGCTACCGGAGACTTCTACACAGGAATATCCCCATCAGTT 202

QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 CCTACCAATGAGTCTGTTGAGGAATAATACATCTAGCATCCTCGGAAGTACAGCCAC 262

QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 TACTTGATCTATCAGGACAGCAGTAAAGCACACACGCCATCTCAGAGACTATGGTCAAC 322

QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 TTTACAGTTACCTCTGGATGCTTCCCTAGGCTCTGGAACTCCACAGCTTTTTCACACCA 382

QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 383 CAGACTTCCCAACTGGCATCTACTTACTTTCAGACAGATATTCACCTTCAGAGATG 442

QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 ACCTGGAAGTCCAGCTCGCCATCTATAAATGTTTCTGATTATTCGCTTAATATAGCAGC 502

QY 141 Leu---AlaThrSerProThrLysProTyr-----ThrSerSerSerProIleLeu 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 503 TTTGAGATGATACATCACCCACCGCCCATATGCTTACACATCTCTCTGCTCCG----- 556

```

```

QY 157 SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 557 AGTGCCATTAAGGAGGAATAATCAATGCTCTGGATCCGAGAGTGGAGTTGGCCAGGCT 616

QY 177 IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 ATCTGCTGGAACTAAGTGAAGCATCTAGTTGTGAGGAGTTTAAAGAGGAAAGGAGAA 676

QY 197 GlyLeuAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnVal 216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 677 GATCTAATTAATACTGCTGAAAGGAGGAGGCTGAGGCTGATGCTGCTGCTAGTGTCT 736

QY 217 CysSerLeuLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 737 TGTCTCCTGCTCTAGCCCACTCTGAGTTAGGCTGAGTGTGCTGATGCTGCTGCTGCT 796

QY 237 AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 797 AATAGCACAGAACTCCAGCAAACTCCAGCTTATGGAAGACCAATCTGACTTGAGA 856

QY 257 LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGln 276
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 857 AAGCTGGGATCCAAAGCTTCAATAACAGATATCGGAGCCACAGAGCTATTCCTCGA 916

QY 277 LysThrLeuIleAlaLeuValThrSerGlyValAlaLeuAlaValLeuGlyIleThrGly 296
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 917 AAGACTCTTATTCATGCTGCTCACTCTGGAGTCTGCTGCCATCTCTGGCACCACCTG 976

QY 297 TyrPheLeuMetAsnArgSerTyrProThrGlyGluArgLeuGluLeuGluPro 316
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 977 TATTCTCTGATGAACGCTCGAGTTCGAGCCCTACAGGAGAAAGGCTGGTGAAGACCT 1036

RESULT 12
AF202879 AF202879 581 bp mRNA linear PRI 02-DEC-1999
LOCUS Homo sapiens hematopoietic progenitor cell antigen CD34 precursor
DEFINITION (CD34) mRNA, partial cds.
ACCESSION AF202879
VERSION AF202879.1 GI:6503195
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 581)
AUTHORS Kutlar,F., Brisko,J., Leithner,C. and Kutlar,A.
TITLE Direct isolation and cDNA sequencing of mRNA from human
hematopoietic progenitor cell antigen CD34
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 581)
AUTHORS Kutlar,F., Brisko,J., Leithner,C. and Kutlar,A.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-1999) Medicine, Hematology and Oncology, Sickle
Cell Center, Medical College of Georgia, 15th Street, AC-1000,
Augusta, GA 30912, USA
FEATURES
Location/Qualifiers
1..581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q32"
<1..>581
/gene="CD34"
<1..>581
/gene="CD34"
/codon_start=2
/product="hematopoietic progenitor cell antigen CD34
precursor"
/db_xref="GI:6503196"
/translation="ILSDIKAEIKSGIREVKLTGICLQNKTSCEAFKKDRSEGL
ARVLCEQADADAGAQVCSLLLAQSEVRPQCLLLVLANRTEISSKLQMKRHSDLK

```


BASE COUNT 150 a 146 c 171 g 114 t
ORIGIN
KLGILDFTQDVASHQSYKSLTIALVTSGALLAVLGTIGTYFLMNRSSWSPGTGERLGE
DPYITENGCGGYSGGPQTSPEACGKASVNRGA"

Alignment Scores:
Pred. No.: 7.09e-45 Length: 581
Score: 793.00 Matches: 159
Percent Similarity: 98.77% Conservative: 1
Best Local Similarity: 98.15% Mismatches: 2
Query Match: 49.84% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-4 (1-316) x AF202879 (1-581)

QY 155 IleLeuSerAspIleLysAlaGluIleLysCysSerGlyValArgGluValLysLeuThr 174
Db 2 ATCTTAAGTGACATCAAGCAGAAATCAAAATGTTACGGCATCAGAGAAATGAAATGACT 61
QY 175 GlnGlyLeuCysLeuGluGlnAsnLysThrSerSerCysAlaGluPhelLysLysAspArg 194
Db 62 CAGGCGATCTGCTGGAGCAAAATAGACCTCAGCTGTGGGAGTTTAAGRAGGACAGG 121
QY 195 GlyGluGlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAlaGlyAla 214
Db 122 GGAGAGGGCTGGCCGAGTGTGTGGGGAGGAGCGCTGATGCTGATGCTGGGGCC 181
QY 215 GlnValCysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuVal 234
Db 182 CAGGTATGCTCCCTGCTCTGCCAGTCTGAGGTGAGGCCCTCAGTGTCTACTGCTGGTC 241
QY 235 LeuAlaAsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAsp 254
Db 242 TTGGCCAAACAGACAGAAATTTCCAGCAAACTCCAACTTATGAAAAGCACCACATCTGAC 301
QY 255 LeuLysLysLeuGlyIleLeuAspPheThrGluGlnAspAlaSerHisGlnSerTyr 274
Db 302 CTGAAAAGCTGGGATCTAGATTTCACTGAGCAAGATGTTGCAAGCCACCAGAGCTAT 361
QY 275 SerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuValLeuGlyIle 294
Db 362 TCCAAAAGACCCGTGATTCACCTGGTCACCTCGGGAGCCCTGCTGGCTGTCTTGGGCATC 421
QY 295 ThrGlyTyrPheLeuMetAsnArgSerTrpSerProThrGlyGluArgLeuGluLeu 314
Db 422 ACTGGCTATTCTGTGATGATCCCGCAGCTGGAGCCCCACAGGAGAAAGGCTGGCGCGAA 481
QY 315 GluPro 316
Db 482 GACCCT 487

RESULT 13
AB021662
LOCUS
DEFINITION Bos taurus mRNA for CD34, complete cds.
ACCESSION AB021662
VERSION AB021662.1 GI:4996448
KEYWORDS CD34.
SOURCE Bos taurus (strain:Holstein) cDNA to mRNA.
ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE
1 (sites)
AUTHORS Zhou,J.H., Hikono,H., Ohta,M. and Sakurai,M.
TITLE Cloning of bovine CD34 cDNA
JOURNAL J. Vet. Med. Sci. 63 (9), 1051-1053 (2001)
MEDLINE 21519980

REFERENCE
2 (bases 1 to 2453)
AUTHORS Sakurai,M., Zhou,J., Hikono,H. and Ohta,M.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1998) Michiharu Sakurai, National Institute of
Animal Health, Laboratory of Immunobiology; 3-1-1 Kannondai,

Tukuba, Ibaraki 305-0856, Japan (E-mail:yukari@nih.affrc.go.jp,
Tel:81-298-7790, Fax:81-298-7880)
FEATURES
source Location/Qualifiers
1..2453
/organism="Bos taurus"
/strain="Holstein"
/db_xref="taxon:9913"
50..1198
/gene="CD34"
50..1198
/gene="CD34"
/note="cluster of differentiation antigen 34"
/codon_start=1
/product="CD34"
/protein_id="BAA78476.1"
/db_xref="GI:4996449"
translation="MLGRRGAGRGMYRGWTALCLLSLPSMETQTPMAATTATVS
ASPEISSPVTTFEDRRNGSSTFRSLDYSDNSGTTAVTLVTPASTDSFTSET
LPASGAVNSVQSOTSLATTVSSSTLISFATSEATLQPSSTCEEFKKNNEKLTGFLC
ITPSFPTQSILKSEIKCRKVKVLAHSICLKNETSCCEEFKKNNEKLTGFLC
OKKQAEVCSLLLAOSEVRPQCLLVLTNRTESSSKIKLLEHMSDLRWMEDISEEA
VSSHQSYSRKTLIALVTSGLLAILITCYFLMNRSSWSPGTGERLGEPPYVENGSGG
GYSQSPASPEAQKAIYNRGAQENGTOATSRNGHSARQVRVADTEL"
BASE COUNT 522 a 708 c 654 g 569 t
ORIGIN

Alignment Scores:
Pred. No.: 1.05e-42 Length: 2453
Score: 772.50 Matches: 179
Percent Similarity: 65.95% Conservative: 36
Best Local Similarity: 54.91% Mismatches: 88
Query Match: 48.55% Indels: 23
DB: 4 Gaps: 5

US-09-836-602-4 (1-316) x AB021662 (1-2453)

QY 1 MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 86 ATGGTGGGGGCTGGACCGGCTCTGCCCTGCTGAGTCTGCTGCTCTCT----- 133
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 134 -----ATGGAATTCACAGACACACCATGGCAGCTACTACTGCCACAGTGTCA---GCC 184
QY 41 SerThrAsnValSerTyrGlnGluThrThr-----Pro 52
Db 185 AGCCCAAGAAATATCTTCACCTGTCCCTACAACTATCGACTGGAGACGATGAAGTCA 244
QY 53 SerThrLeuGlySerThrSerLeuHisProValSerGln-----HisGlyAsnGluAla 70
Db 245 AGTACTTTTAGAGACACACAGCCCTCTACGATGCTCTCAGGACAGCAATGGGACCACAGCA 304
QY 71 ThrThrAsnIleThrGluThrValLysPheThrSerThrSerValIleThrSerVal 90
Db 305 GTCACCTTAGTTCACGCGCTACAGACAGTTTCCACGTCTACCTCTGAGACCCCTCCAGCC 364
QY 91 TyrGlyAsnThrAsnSerSerValGlnSerGlnThrSerValIleSerThrValPheThr 110
Db 365 TCTGGGGCCGTGAACCTCTTCTGCCAGTCACAGACCTCTTACCCACCGGCTCTTCT 424
QY 111 ThrProAlaAsnValSerThrProGluThrThrLeuLysProSerLeuSerProGlyAsn 130
Db 425 ACCCTCATCAGCTTCGGACTTCAGAGCGGACCTCTGCAGCCACGACTTTCTCTGGAAT 484
QY 131 ValSerAspLeuSerThrThrSerThrSerLeuAlaThrSerProThrLysProTyrThr 150
Db 485 ATTTGGATTCCCTGTACAATAGTACCAGCCCTGTGACTTCTTCCATTATTACTCTCCA 544
QY 151 SerSerSerProIleLeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGlu 170
Db 545 TCATTTTCTCCTACCCAAAGTATCCTCAAGAGTGAATAAAATGTTCCAGAGTCAAGAA 604
QY 171 ValLysLeuThrGlnGlyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPhe 190

Contains CD34 (hematopoietic stem cell surface antigen), CA repeat, EST, GSS, complete sequence.

ACCESSION
AL035091
VERSION
KEYWORDS
SOURCE

AL035091.2 GI:4581417
HTG; CD34; repeat polymorphism.
human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 160771)
Graham.D.

REFERENCE

AUTHORS
TITLE
JOURNAL
Submitted (22-FEB-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

COMMENT

On Apr 12, 1999 this sequence version replaced gi:4140364.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence is the entire insert of clone 88L2.
The true right end of clone 7H11 is at 71906 in this sequence. 8L2 is from the library RCI6 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:ppAC4>.

FEATURES

source

1. .160771
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"

/map="q32.2-q32.3"
/clone="XX-D88L2"
/clone.lib="RPC16"

repeat_region

42. .81

/note="10 copies 4 mer aaga 83% conserved"

42. .75

/note="17 copies 2 mer aa 88% conserved"

89. .412

/note="L1MC5 repeat: matches 7609. .7913 of consensus"

1027. .1154

/note="L2 repeat: matches 2586. .2744 of consensus"

1189. .1277

/note="L2 repeat: matches 2616. .2704 of consensus"

2426. .2643

/note="MER58A repeat: matches 1. .224 of consensus"

complement(3075..3399)

/note="match: 252805 STS containing (CA) repeat"

3121. .3176

/note="28 copies 2 mer gt 100% conserved; differs from

252805"

3806. .3905

/note="L2 repeat: matches 2600. .2750 of consensus"

8052. .8339

/note="L2 repeat: matches 1823. .2111 of consensus"

8406. .8600

/note="MLT1A1 repeat: matches 162. .365 of consensus"

8631. .8815

/note="L2 repeat: matches 2158. .2346 of consensus"

8843. .9132

/note="AluJb repeat: matches 1. .292 of consensus"

9295. .9363

misc_feature

9372. .9672

/note="match: GSS AQ075766 clone 2359015"

repeat_region

9493. .9594

/note="MIR repeat: matches 96. .205 of consensus"

repeat_region

9619. .9640

/note="11 copies 2 mer tc 100% conserved"

repeat_region

9671. .9835

/note="MIR repeat: matches 46. .200 of consensus"

repeat_region

10253. .10285

/note="L2 repeat: matches 2708. .2740 of consensus"

repeat_region

10284. .10406

/note="MIR repeat: matches 64. .196 of consensus"

prim_transcript

<11347. .11584

/note="match: multiple ESTs; match: AA884400 AA971811"

repeat_region

11846. .11878

/note="MIR repeat: matches 1. .33 of consensus"

repeat_region

11879. .12182

/note="AluJ repeat: matches 1. .303 of consensus"

repeat_region

12183. .12411

/note="MIR repeat: matches 33. .258 of consensus"

repeat_region

12443. .12491

/note="L2 repeat: matches 2662. .2710 of consensus"

repeat_region

12574. .12661

/note="MIR repeat: matches 178. .262 of consensus"

repeat_region

12756. .13049

/note="AluJb repeat: matches 1. .300 of consensus"

misc_feature

13231. >13541

/note="match: STS G28100, G24984"

prim_transcript

13231. .514597

/note="match: multiple ESTs; match: 3' EST T89361 clone

110082; Paired with EST T84126 matching this clone; match:

5' EST T84126 clone 110082; Paired with EST T89361

matching this clone"

repeat_region

13672. .13875

/note="MIR repeat: matches 29. .231 of consensus"

repeat_region

14598. .14907

/note="AluJ repeat: matches 1. .310 of consensus"

repeat_region

14913. .15214

/note="AluJb repeat: matches 7. .294 of consensus"

repeat_region

15376. .15596

/note="L2 repeat: matches 1263. .1527 of consensus"

repeat_region

15606. .15666

/note="MLT1F repeat: matches 478. .541 of consensus"

repeat_region

15667. .16045

/note="L1PA3 repeat: matches 5764. .6146 of consensus"

repeat_region

16046. .16084

/note="MLT1F repeat: matches 438. .478 of consensus"

repeat_region

16085. .16379

/note="AluJb repeat: matches 1. .296 of consensus"

repeat_region

16380. .16850

/note="MLT1F repeat: matches 1. .438 of consensus"

repeat_region

16854. .17208

/note="L2 repeat: matches 1533. .1930 of consensus"

repeat_region

17268. .17395

/note="L1MC4 repeat: matches 7854. .7977 of consensus"

repeat_region

17422. .17485

/note="Charliel repeat: matches 79. .143 of consensus"

repeat_region

17509. .17625

/note="L1MC4 repeat: matches 7706. .7822 of consensus"

repeat_region

17627. .17677

/note="L2 repeat: matches 1949. .1995 of consensus"

repeat_region

17678. .17988

/note="AluJ repeat: matches 1. .311 of consensus"

repeat_region

17989. .18227

/note="L2 repeat: matches 1995. .2257 of consensus"

repeat_region

18250. .18508

/note="Charlie5 repeat: matches 2315. .2577 of consensus"

repeat_region

18510. .18698

/note="Charlie5 repeat: matches 16. .196 of consensus"

repeat_region

18721. .18935

/note="Charlielb repeat: matches 3. .219 of consensus"

repeat_region

18936. .19243

repeat_region /note="AluY repeat: matches 1. .307 of consensus" 19244. .19531
repeat_region /note="Charlieb repeat: matches 219. .501 of consensus" 19610. .19779
repeat_region /note="L1MD repeat: matches 1. .169 of consensus" 19780. .20228
repeat_region /note="L1MD repeat: matches 787. .1293 of consensus" 20299. .20606
repeat_region /note="L2 repeat: matches 2382. .2710 of consensus" 20625. .20753
repeat_region /note="L2 repeat: matches 2179. .2307 of consensus" 20813. .20960
repeat_region /note="L2 repeat: matches 2374. .2523 of consensus" 21149. .21292
repeat_region /note="MIR repeat: matches 110. .261 of consensus" 21334. .22002
repeat_region /note="L2 repeat: matches 1947. .2724 of consensus" <21938. .22360
misc_feature /note="match: GSS AQ193501 clone 2383C13" <22003. .>22108
prlm_transcript /note="match: multiple ESTs; match: AA909597 AA906144 AA907722"
repeat_region 22573. .22657
misc_feature /note="MIR repeat: matches 81. .160 of consensus" complement(24084. .25666)
mRNA /note="match: STS G06650" complement(join(24102. .25486,26245. .26409,26724. .26776, 27028. .27184,35057. .35137,36529. .36782,37377. .37559, 48563. .48899))
gene /gene="CD34"
/note="match: M81104; x60172; match: multiple ESTs; match: T80274 C18678 AA553071 AA311526 AA311473 AA907722 AA906144 AA064307 A1173145 AA369014 R39493 AA148360 T81557 R72478 R69703 R69740 AA039660 RA639555 R73090 R73955 W92532 R66845 AA054965 AA434483 R69214 R82715 AA188435 AA524276 W7409 H01096 T12007 AA216274 AA022916 AA528783 AA483672 AA640408 W58493 T28504 A1128488 A1149563 A1144193 A1252665 A1017002 A1160509 A1017793 AA906021 A1128776 A1150834 AA434387 AA664247 AA039661 W72884 A1262206"
evidence=not_experimental
complement(24102. .48899)
/gene="CD34"
CDS complement(join(25301. .25486,26245. .26409,26724. .26776, 27028. .27184,35057. .35137,36529. .36782,37377. .37559, 48563. .48605))
/gene="CD34"
/note="match: SWISS-PROT:P28906"
/codon_start=1
evidence=not_experimental
/product="CD34 (hematopoietic stem cell surface antigen)"
/protein_id="CAB36556.1"
/db_xref="GI:4455631"

Alignment Scores:

Pred. No.: 5,35e-30 Length: 160771
Score: 617.00 Matches: 145
Percent Similarity: 42.32% Conservativity: 1
Best Local Similarity: 42.03% Mismatches: 1
Query Match: 38.78% Indels: 198
DB: 9 Gaps: 1

US-09-836-602-4 (1-316) x HS88L2 (1-160771)

QY 14 LeuProSerGlyPheMetSerLeuAsnAsnGlyThrAlaThrProGluLeuProThr 33
Db 37563 ATAGCTTCGGGTTTCATGAGTCTTGACAAACACGGTACGTGACCCCAAGTACTACTAC 37504

QY 34 GlnGlyThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSer 53
Db 37503 CAGGGAACATTTTCAAATGTTTCTACAAATGATCCTACCAAGAAACATACACACCTAGT 37444

QY 54 ThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsn 73
|||||

Search completed: October 30, 2002, 09:52:02
Job time : 1666.9 secs

Db 37443 ACCCTTGGGAAGTACCAGCCTGCACCCTGTGTCTCAACATGGCAATGAGGCCACACAAC 37384
QY 74 IleThr----- 75
|||||
Db 37383 ATCACAGGTAAACACAGCATTTTGTCTCAGATCCCGGAGAGATGCTGGTGATGCTTGGGTA 37324
QY 75 ----- 75
Db 37323 AAGCATTTAGGATGTTTTCAGACCGCTCCCTCTCCACAGAGAAATTTATACAAGTCCCT 37264
QY 75 ----- 75
Db 37263 AGTATTAAATGACTTGGATATCATGCTTAGGGTGCCCTGAAGTAAGTCTTAGATAATCTTT 37204
QY 75 ----- 75
Db 37203 CCTTCAGTCACAGTTTCTTCAGGCCCATGCTTTGGGAGCTGATCCTCGATCAGTAATGCCT 37144
QY 75 ----- 75
Db 37143 CCATACCCCTTCTCTCATGCTGAGTCTAGTCTAGTATAAAATAAGGATCAGAAGACTCA 37084
QY 75 ----- 75
Db 37083 ATAGAGATCTCCGATCTCTTTTAAAGAAAGAGGTGGGTACAGGTCCAAACTGGGGTC 37024
QY 75 ----- 75
Db 37023 TTTGGCTTCTCTAAGGTAGACCAGCATCTATTTCAGTTTCAGGACACCTCTAATACATTC 36964
QY 75 ----- 75
Db 36963 TAGATTCTAGCTCTGTGTTCCACAGCTTGAAATGAGTTTGGTCAGGGATGGACACAGAG 36904
QY 75 ----- 75
Db 36903 TAACTGTTAACTCCCAAACTCCCTTTTCTGTGTTGAAGTGCATGCCATATCATGACCT 36844
QY 75 ----- 75
Db 36843 GAGATTTTGTGTATCTATGAGCAGCATGCACATGCACACCATGTTTGGTCTCTTCCA 36784
QY 76 GluThrThrValLysPheThrSerThrSerValIleThrSerValIleThrSerValIleThrAsn 95
|||||
Db 36783 GAAACGACAGTCAAATTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACACAAC 36724
QY 96 SerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrProAlaAsnVal 115
|||||
Db 36723 TCTTCTGTCCAGTCACAGACCTCTGTAATFCAGCAGAGTTTCACCCACCCACCCACGTT 36664
QY 116 SerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSer 135
|||||
Db 36663 TCACTCCAGACACAACCTTGAGGCTAGCCCTGTACCTGGAAATGTTTTCACACCTTCA 36604
QY 136 ThrThrSerThrSerLeuAlaThrSerProThrLysProTyrThrSerSerSerProIle 155
|||||
Db 36603 ACCACTAGCACATGCTTGCAACATCTCCCACTAAACCCCTATACATCATCTCTCTCTATC 36544
QY 156 LeuSerAspIleLys 160
|||||
Db 36543 CTAAGTGACATCAAG 36529

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:10:53 ; Search time 1294.57 Seconds
(without alignments)
3888.820 Million cell updates/sec

Title: US-09-836-602-2

Perfect score: 1895

Sequence: 1 MPRGWTALCLLLPLSGFMS.....QATSRNGHSARQHVADTEL 373

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlip
-Q=/cgn2_1/USPFO_Spool/US09836602/runat_29102002_101144_3812/app_query.fasta.1.1429
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09836602 -CGN_1_1_2453_@runat_29102002_101144_3812 -NCPU=6 -ICPU=3
-NO_XLIP -NO_MMAB -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pla:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1548	81.7	977	9	AL548050
2	1409	74.4	919	9	AL513812

3	1153	60.8	735	9	AL598821
4	1122	59.2	761	9	AU141253
5	1096	57.8	752	9	AU141111
6	1078	56.9	749	9	AU121093
7	1069.5	56.4	768	9	AU141883
8	1045	55.1	1129	10	BI769224
9	994	52.5	804	10	BG675028
10	991.5	52.3	754	9	AU135182
11	907.5	47.9	1001	10	BG256232
12	792	41.8	717	9	AU121843
13	791	41.7	997	9	AU547847
14	755.5	39.9	733	9	AU141125
15	714	37.7	825	9	AU139201
16	700	36.9	839	10	BF178239
17	685.5	36.2	903	10	BI824944
18	670	35.4	780	9	AI806138
19	627	33.1	464	10	BM146840
20	627	33.1	551	10	T80274
21	625	33.0	549	10	BE948643
22	624	32.9	477	10	BM146845
23	611	32.2	558	9	AM656254
24	571.5	30.2	603	9	AA673625
25	561	29.6	473	10	BF830285
26	555	29.3	486	9	AI893233
27	553	29.2	657	9	AL547643
28	541	28.5	382	10	C18678
29	538	28.4	748	10	BI152175
30	536	28.3	529	10	BM069265
31	533	28.1	464	9	AA792789
32	529	27.9	424	10	W65699
33	521	27.5	672	10	BI692649
34	498	26.3	518	9	BE236076
35	477.5	25.2	711	10	BF214180
36	467	24.6	477	9	AI371138
37	458	24.2	306	9	AI595477
38	446.5	23.6	397	10	BF993687
39	423	22.3	520	9	AA064307
40	416	22.0	805	10	BG864235
41	411	21.7	806	10	BE913124
42	404.5	21.3	1041	10	BF139664
43	365	19.3	223	10	H72215
44	361.5	19.1	675	9	BB665591
45	342.5	18.1	575	9	AA907722

ALIGNMENTS

RESULT 1	AL548050	977 bp	mRNA	linear	EST 16-FEB-2001
LOCUS	AL548050	LTI_NFL006_PL2	Homo sapiens	cdna clone	CS0D1034YP22 5
DEFINITION	AL548050	LTI_NFL006_PL2	Homo sapiens	cdna clone	CS0D1034YP22 5
ACCESSION	AL548050	GI:12882692			
VERSION	AL548050.1	GI:12882692			
KEYWORDS	EST				
SOURCE	human				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.				
TITLE	Full-length cDNA libraries and normalization				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.				
FEATURES	Location/Qualifiers				
Source	1..977				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="CS0D1034YP22"				
	/clone_lib="LTI_NFL006_PL2"				

/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 259 a 272 c 231 g 214 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 5.18e-123 Length: 977
Score: 1548.00 Matches: 312
Percent Similarity: 99.68% Conservative: 0
Best Local Similarity: 99.68% Mismatches: 1
Query Match: 81.69% Indels: 1
DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x AL548050 (1-977)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 40 ATGCCGGGGCTGGAGCGCGCTTGTGCTGAGTTGCTGCTTCTGGGTTTCATGAGT 99
QY 21 LeuAspAsnGlnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 100 CTTGACAAACACGGTACTGTACCCAGAGTTACCTACCCAGGGAACATTTTCAAAATGTT 159
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 160 TCTACAATGTATCTACCAAGAACTACACACCTAGTACCTTGGAGTACAGCGCTG 219
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
Db 220 CACCCTGTGTCTCAACATGGCAATGAGGCCACAAACATCACAGAAACGACAGTCAAA 279
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 280 TTCACATCTACCTCTGTGTAACCTCAGTTTATGGAAACACAACTCTTCTGTCAGTCA 339
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 340 CAGACCTCTGTATCAGCAGAGTTCCACCCAGCCAGCTTCACTCCAGAGACA 399
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 400 ACCTTGAAGCCTAGCCCTGTACCTGGAAATGTTTCAGACCTTTCAACCCACTAGCCTAGC 459
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLys 160
Db 460 CTTGCAACATCTCCCACTAAACCTATACATCATCTCTCTATCTTAAGTGACATCAAG 519
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 520 GCAGAAATCAAAATGTTCCAGGCATCAGAGAACTGAATGTGACTCAGGGCATCTCGCTGGAG 579
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 580 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAGACAGGGAGGCGCTGGCCCGA 639
QY 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 640 GTCTGTGTGGGAGGAGCAGGCTGATGCTGATGCTGGGGCCAGGATGCTCCCTGCTC 699
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 700 CTTGCCCCAGTCTCAGGTGAGGCCCTCAGTGTCTACTGCTGTGCTTGGCCACAGACAGAA 759
QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260

Db 760 ATTTCCAGCAATCCCAACTTATGAAAAGCACCACAACTCTGACCTGAAAAGCTGGGATC 819
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 820 CTAGATTTCACCTAGCAGAGATGTGCAAGCCACAGAGCTATTCCCAAGACCTGATT 879
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 880 GCACGTGTCACCTCGGAGCCCTGCTGGCTGTCTTGGGCATCACTGGCTATTTCCTCATG 939
QY 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGly 313
Db 940 AATCGCGGAGCTGGAGCCCMCA-GGAGAAAGGCTGGGG 977
RESULT 2
AL513812 919 bp mRNA linear EST 13-FEB-2001
LOCUS AL513812 LTI_NFL006_PL2 Homo sapiens cDNA clone CL0BA0092F10 5
DEFINITION prime, mRNA sequence.
ACCESSION AL513812
VERSION AL513812.1 GI:12777306
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 919)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source

1. 919
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CL0BA0092F10"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 246 a 253 c 213 g 198 t 9 others
ORIGIN

Alignment Scores:
Pred. No.: 4.22e-111 Length: 919
Score: 1409.00 Matches: 286
Percent Similarity: 97.61% Conservative: 0
Best Local Similarity: 97.61% Mismatches: 7
Query Match: 74.35% Indels: 1
DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x AL513812 (1-919)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 40 ATGCCGGGGCTGGAGCGCGCTTGTGCTGAGTTGCTGCTTCTGGGTTTCATGAGT 99
QY 21 LeuAspAsnGlnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 100 CTTGACAAACACGGTACTGCKACCCAGAGTTACCTACCCAGGGAACATTTTCAAAATGTT 159

QY	41	SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu	60
DB	160	TCACAAATGATATCCYACCAAGAACTACAACTAGTACCTTGGAAAGTACAGCCTG	219
QY	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys	80
DB	220	CACCTGTGTCTCAACATGGCATGAGGCCACAACAACATCACAGAAACACAGTAA	279
QY	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer	100
DB	280	TTACATCTACCTCTGTGATACCTCAGTTATATGAAACACAACTCTCTGTCCAGTCA	339
QY	101	GlnThrSerValIleSerThrValPheThrThrProLaaSnValSerThrProGluThr	120
DB	340	CAGACCTCTCTAATCAGCAGAGTTTCCACCACCCAGCCAACTGTTCAACTCCAGAGACA	399
QY	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrSerThrSer	140
DB	400	ACCTTGAAGCTAGCTGTCTCACCTGGAAATGTTTCAGACCTTTTCACCACTAGCAGTACG	459
QY	141	LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys	160
DB	460	CTTGCAACATCTCCCACTAAACCCATATACATCATCTCTCTATCCYAAAGTACATCAAG	519
QY	161	AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu	180
DB	520	GCAGAAATCAAAATGTTTCAGGCATCAGAGAAAGTAAATGACTCAGGGCATCTGCCTGGAG	579
QY	181	GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg	200
DB	580	CAAAATGAAGACTCCACCTGTGCGGAGTTTAAAGAGACAGGGAGAGGGCTGGCCGA	639
QY	201	ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu	220
DB	640	GTGCTGTGTGGGAGGAGCAGGCTGATGCTGATGCTGGGGCCAGGTATGCTCCCTGCCTC	699
QY	221	LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu	240
DB	700	CTTGCCCAAGCTGAGGTGAGGCCCTCACTGTCTACTGTGTGCTTGGCCAAACAGACAGAA	759
QY	241	IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle	260
DB	760	ATTTCCAGCAACTCCCACTTATGAAAGACCACTGACCTGACCAAGCTGGGATC	819
QY	261	LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle	280
DB	820	YTAGATTTCAGTGAGCMAGAWTTGCAAGCCACAGACCTATTTCCCAAAAGACCTGATT	879
QY	281	AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGly	293
DB	880	GCAGTGTCACTCGGGAGC-CTGCTGGGTGCTTGGGC	917
RESULT 3			
AL598821	735 bp	mRNA	linear EST 14-AUG-2001
LOCUS	DKFZp313K2322_r1 313 (synonym: hlcc2)	Homo sapiens	cdNA clone
DEFINITION	DKFZp313K2322 5', mRNA sequence.		
ACCESSION	AL598821		
VERSION	AL598821.1	GI:15161512	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 735)		
	Duesterhoef,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann		
	,S.		
TITLE	EST (Duesterhoef, et al.)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Duesterhoef A		
	MIPS		
	Am Klopfererspitz 18a D-82152 Martinsried, Germany		
	This is the 5' sequence of the clone insert		

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Olagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.

No sl sequence available.

This clone (DKFZp3K2322) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY. Email: clone@rzpd.de.

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10
11. Feature 11	Source 11
12. Feature 12	Source 12
13. Feature 13	Source 13
14. Feature 14	Source 14
15. Feature 15	Source 15
16. Feature 16	Source 16
17. Feature 17	Source 17
18. Feature 18	Source 18
19. Feature 19	Source 19
20. Feature 20	Source 20
21. Feature 21	Source 21
22. Feature 22	Source 22
23. Feature 23	Source 23
24. Feature 24	Source 24
25. Feature 25	Source 25
26. Feature 26	Source 26
27. Feature 27	Source 27
28. Feature 28	Source 28
29. Feature 29	Source 29
30. Feature 30	Source 30
31. Feature 31	Source 31
32. Feature 32	Source 32
33. Feature 33	Source 33
34. Feature 34	Source 34
35. Feature 35	Source 35
36. Feature 36	Source 36
37. Feature 37	Source 37
38. Feature 38	Source 38
39. Feature 39	Source 39
40. Feature 40	Source 40
41. Feature 41	Source 41
42. Feature 42	Source 42
43. Feature 43	Source 43
44. Feature 44	Source 44
45. Feature 45	Source 45
46. Feature 46	Source 46
47. Feature 47	Source 47
48. Feature 48	Source 48
49. Feature 49	Source 49
50. Feature 50	Source 50
51. Feature 51	Source 51
52. Feature 52	Source 52
53. Feature 53	Source 53
54. Feature 54	Source 54
55. Feature 55	Source 55
56. Feature 56	Source 56
57. Feature 57	Source 57
58. Feature 58	Source 58
59. Feature 59	Source 59
60. Feature 60	Source 60
61. Feature 61	Source 61
62. Feature 62	Source 62
63. Feature 63	Source 63
64. Feature 64	Source 64
65. Feature 65	Source 65
66. Feature 66	Source 66
67. Feature 67	Source 67
68. Feature 68	Source 68
69. Feature 69	Source 69
70. Feature 70	Source 70
71. Feature 71	Source 71
72. Feature 72	Source 72
73. Feature 73	Source 73
74. Feature 74	Source 74
75. Feature 75	Source 75
76. Feature 76	Source 76
77. Feature 77	Source 77
78. Feature 78	Source 78
79. Feature 79	Source 79
80. Feature 80	Source 80
81. Feature 81	Source 81
82. Feature 82	Source 82
83. Feature 83	Source 83
84. Feature 84	Source 84
85. Feature 85	Source 85
86. Feature 86	Source 86
87. Feature 87	Source 87
88. Feature 88	Source 88
89. Feature 89	Source 89
90. Feature 90	Source 90
91. Feature 91	Source 91
92. Feature 92	Source 92
93. Feature 93	Source 93
94. Feature 94	Source 94
95. Feature 95	Source 95
96. Feature 96	Source 96
97. Feature 97	Source 97
98. Feature 98	Source 98
99. Feature 99	Source 99
100. Feature 100	Source 100

```

1.:733
/organism="Homo sapiens"
/ab_xref="taxid:9606"
/clone_lib="3132322"
/clone_lib_key="3132322"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplex2; Site_1: SfiIA;
cdna-collection"

```

BASE COUNT	CDNA COLLECTION	2 others
ORIGIN		
191 a	205 c	164 t

Alignment Scores:		
Pred. No.:	3 1e-89	Length: 735
Score:	153.00	Matches: 233
Percent Similarity:	98.31%	Conservative: 0
Best Local Similarity:	98.31%	Mismatches: 2
Query Match:	60.84%	Indels: 2
DB:	9	Gaps: 0

US-09-836-602-2 (1-373) x AL598821 (1-735)

Qv 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20

Db 28 ATGCCCGGGGCTGGACCGCGCTTTGCTTGCTGAGTTTGCTGCCCTTCTGGGTTTCATGAGT 87

21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40

D_b 88 CTTGACACAAACGGTACTGGTACCCGAGAGTTACCTACCCAGGGAACAATTTCAAAATGTT 143

0v 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlyValSerThrSerLeu 60

db 148 TCTACAAATGTTTCTCCCAACAACCTACTACCCTGGCAACTACCTTC

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible]

Q4 LEUATATIM SEIFTOIM LYSFOIYIIM SEISEFIOITREUSEIAPRIEYS IO

DB 448 CTTGCAACATCTCCCACTAAACCCATATACATCATCTCTCTATCTCCTAAAGTGACATCAAG 30

QY I6I AIAGIUIIeLYscysseRGLYIleARfGluVAlLYsLeuThfGInGLYIleCYsLeuGLU I8

D6 508 GCAGAAATCAAATGTTTCAGGCATCAGAGAAGTGAAATTGACTCAGGGCATCTGCCCTGGAG 56

QY 181 GlnAsnLysThrSerSerCysAlaGluPheLys-LysAspArg-GlyGluGlyLeuAlaA 200

Db 568 CAAATAAGACCTCCAGCTGTCGGAGTTAAGAAAGGACAGGGGAGAGGGCCTGGCCC 62

Qy 200 rgValLeuCysGlyGluGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuL 2200

.....

[illegible]

Score:	1096.00	Matches:	218
Percent Similarity:	98.20%	Conservative:	0
Best Local Similarity:	98.20%	Mismatches:	4
Query Match:	57.84%	Indels:	0
Db:	9	Gaps:	0
US-09-836-602-2 (1-373) x AU141111 (1-752)			
Qy	1	MetProArgGlyThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer	20
Db	65	ATGCGCGGGCTGACCGCTTTCGTGCTGAGTTTGCCTTCCTGGGTTTCATGAGT	124
Qy	21	LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
Db	125	CTTGACAACAAGGTAAGTACCCAGAGTTACCTACCCAGGGAACATTTCAATGTT	184
Qy	41	SerThrAsnValSerTyrglnGluThrThrProSerThrLeuGlySerThrSerLeu	60
Db	185	TCTACAAATGTATCTACCAAGAACTACAACACCTAGTACCTTTGGAGTACCAAGCCTG	244
Qy	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys	80
Db	245	CACCTGTGTCTCAACATGGCAATGAGGCCCAACAACATCACAGAACGACGTCAA	304
Qy	81	PheThrSerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValGlnSer	100
Db	305	TTACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAACTCTCTGCCAGTCA	364
Qy	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	365	CAGACCTCTGTAATCAGCACAGTGTACCCACCCAGCCCAAGTTTCAACTCCAGAGACA	424
Qy	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer	140
Db	425	ACCTTGAGCCCTAGCCTGTCACCTGGAATGTTTCAGACCTTTCACCACTAGCACTAGC	484
Qy	141	LeuAlaThrSerProThrLysProTyrrSerSerSerProIleLeuSerAspIleLys	160
Db	485	CTTGCAACATCTCCCACTAAACCCCTATACATCATCTCTCTCTATCTTAAGTGACATCAAG	544
Qy	161	AlaGluIleLysCysSerGlyIleargGluValLysLeuThrGlnGlyIleCysLeuGlu	180
Db	545	GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCAGGGCATCTGCCCTGGAG	604
Qy	181	GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyClyGluIleLeuAlaArg	200
Db	605	CAAAATAGACCTTCAGCTGTCGGAGTTTAAGAAGGACAGGGGAGAGGGCCTGGCCCCA	664
Qy	201	ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu	220
Db	665	GTGCTGTGTGGGAGGAGCAGCTGATGCTGATGTGGGGCCCAAGTATGCTCTGCTC	724
Qy	221	LeuAla 222	
Db	725	CTTGCC 730	
RESULT	6		
LOCUS	AU121093	749 bp	linear
DEFINITION	AU121093 HEMBB1 Homo sapiens cDNA clone HEMBB1002065 5', mRNA		
ACCESSION	AU121093		
VERSION	AU121093.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	1 (bases 1 to 749)		
	Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,		
	Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and		
	Isogai,T.		
TITLE	HRI human cDNA project		

JOURNAL	Unpublished (2000)
COMMENT	Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3951 Fax: 81-438-52-3952 Email: genomics@hri.co.jp HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.
FEATURES	Location/Qualifiers
source	1..749 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="HEMBB1002065" /clone_lib="HEMBB1" /tissue_type="whole embryo, mainly body" /dev_stage="embryo, 10 weeks" /note="vector: pME18SFL3"
BASE COUNT	195 a 211 c 173 g 166 t
ORIGIN	4 others
Alignment Scores:	
Pred. No.:	8.93e-83
Score:	1078.00
Percent Similarity:	96.83%
Best Local Similarity:	96.83%
Query Match:	56.89%
DB:	9
US-09-836-602-2 (1-373) x AU121093 (1-749)	
Qy	1 MetProArgGlyThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db	73 ATGCCCGGGGCTGGACCGCGCTTTCGTGCTGAGTTTGCCTTCCTGGGTTTCATGAGT 132
Qy	21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db	133 CTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTCAAAATGTT 192
Qy	41 SerThrAsnValSerTyrglnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db	193 TCTACAAATGTATCTACCAAGAAACTACAACACCTAGTACCTCCCTTGAAGTACCAAGCCTG 252
Qy	61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db	253 CACCTGTGTCTCAACATGGCAATGAGGCCCAACAACATCACAGAAACGACAGTCAAA 312
Qy	81 PheThrSerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValGlnSer 100
Db	313 TTCATATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTCTGTCAGTCA 372
Qy	101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db	373 CAGACCTCTGTAATCAGCACAGTGTTCACCAACCCAGCCCAAGCTTTCAACTCCAGAGACA 432
Qy	121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db	433 ACCTTTGAAGCCTAGCCCGTCACTCGGAATGTTTCAGACCTTTTCACCACTATGCACTAGC 492
Qy	141 LeuAlaThrSerProThrLysProTyrrThrSerSerSerProIleLeuSerAspIleLys 160
Db	493 CTTGCAACATCTCCCACTAAACCCCTATACATCATCTTCTCTCTATCTCTAAAGTACATCAAG 552
Qy	161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db	553 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCANGGCTGCTGGGAG 612
Qy	181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db	613 CAAAATAGACCTTTCAGCTGTGGGAGTTTAAGAAGGACAGGGGAGGCGCTGGCCCCA 672

```

QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
|||||
Db 673 GTGCTGTGGGGAGAACANGCTGATGCTGATGCTGGGGCCCAAGTATGCTCTGCTCC 732
QY 221 Leu 221
|||
Db 733 TTG 735

RESULT 7
AUI41883 768 bp mRNA linear EST 25-OCT-2000
LOCUS AUI41883 THYR01 Homo sapiens cDNA clone THYR01001399 5', mRNA
DEFINITION sequence.
ACCESSION AUI41883
VERSION AUI41883
KEYWORDS EST.
SOURCE AUI41883.1 GI:11003404
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 768)
AUTHORS Oka,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamanoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1..768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYR01001399"
/clone_lib="THYR01"
/tissue_type="thyroid gland"
/notes="Vector: pME18SFL3"
BASE COUNT 198 a 215 c 181 g 171 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 4.98e-82 Length: 768
Score: 1069.50 Matches: 217
Percent Similarity: 95.61% Conservatives: 1
Best Local Similarity: 95.18% Mismatches: 10
Query Match: 56.44% Indels: 1
DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x AUI41883 (1-768)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
|||||
Db 72 ATGCCGGGGGCTGGACCGGCTTGTGCTGAGTTTGCTGCTTGTGGGTTTCATGAGT 131
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
|||||
Db 132 CTGTACAAACACGGTACTGTACCCAGAGTTACCTACCCAGGGAACATTTCAAATGTT 191
QY 41 SerThrAsnValSerTyrglnGluThrThrProSerThrLeuGlySerThrSerLeu 60
|||||
Db 192 TCTACAAATGATCTACCAAGAACTACAACTAGTACCTTGGAGTACGAGCTG 251
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
|||||

```

```

Db 252 CACCTGTGCTCAACATGGCAATGAGCCACAAACACACAGAAACACAGTCAAA 311
QY 81 PheThrSerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValGlnSer 100
|||||
Db 312 TTCACTCTACCTCTGTGTAACCTCAGTTTATGGAACACAAACTCTCTGTGTCAGTCA 371
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
|||||
Db 372 CAGACCTCTGTAATCAGCACAGTGTTCACCACCCAGCCAGCGTTTCAACTCCAGAGACA 431
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
|||||
Db 432 ACCTTGAAGCGCTACGCTGTACCTGGAATGTTTCAGACCTTTCACCACTAGCACTAGC 491
QY 141 LeuAlaThrSerProThrLysProTyrrThrSerSerSerProIleLeuSerAspIleLys 160
|||||
Db 492 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCCTAAGTGACATCAAG 551
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
|||||
Db 552 GCAGAAATCAAAATGTTTCAGGCATCAGAGAGTGAATTGACTCAGGGCATCTGCCTGGAG 611
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
|||||
Db 612 CAAAATAAGACCTTCAGCTGTGCGGAGTTTAAAGAAGACAGAGGGAGGCGCTGGCCGA 671
QY 201 ValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
|||||
Db 672 GTGCTGTGGGGAGGAACANGCTGATGCTGCTGGGC--CCAGTATGCTTCTCTCTC 729
QY 221 LeuAlaGlnSerGluValArgPro 228
|||||
Db 730 CTTGCCAATCTGANGTGAAGSCCT 753

RESULT 8
BI769224 1129 bp mRNA linear EST 25-SEP-2001
LOCUS 603060051F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5209333 5',
DEFINITION mRNA sequence.
ACCESSION BI769224
VERSION BI769224.1 GI:15760802
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1129)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11525 row: e column: 14
High quality sequence stop: 765.
FEATURES
Location/Qualifiers
1..1129
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5209333"
/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/notes="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site:1: NotI; Site:2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed

```

upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH_MGC Library.*

BASE COUNT 292 a 317 c 279 g 241 t
ORIGIN

Alignment Scores:
Pred. No.: 1,11e-79 Length: 1129
Score: 1045.00 Matches: 255
Percent Similarity: 77.56% Conservative: 18
Best Local Similarity: 72.44% Mismatches: 46
Query Match: 55.15% Indels: 33
DB: 10 Gaps: 9

US-09-836-602-2 (1-373) x B1769224 (1-1129)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 55 ATGCCGCGGGCTGGACCGCGCTTGTCTGCTGAGTTGCTGCCTTCGGGTTCATGAGT 114
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 115 CTGACAAACACGGTACTGCTACCCAGAGTACCTACCCAGGGAACATTTTCAAATGTT 174
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 175 TCTACAAATGATCTCTACCAAGAACTACAAACCTAGTACCTTGGGAAGTACCAAGCTG 234
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 235 CACCCTGTGTCTCAACATGGCAATGAGCCACACAAACATCACAGAAACGACAGTCAA 294
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 295 TTCACATCTAGCTGTGTATACCTCAGTTTATGGAACACAAACCTCTCTGTCAGTCA 354
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 355 CAGACCTCTGTAATCAGCACAGTGTACCAACCCAGCCAGCAAGTTCAACTCCAGAGACA 414
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 415 ACCITGAAGCCTAGCTGTACCTGGAATGTTTCAGACCTTCAACACACTAGCAGTAC 474
Qy 141 LeuAlaThrSerProThr-LysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 475 CTTCACACATCTCCACCTAAAGCCCTATACATCATCTTCTCTATCTCTAAGTGACATCAA 534
Qy 160 sAlaGluIleLysCysSerGlyIle-ArgGluVal-LysLeuThrGlnGlyIleCysLeu 179
Db 535 GCGAATCAATCAATGTTCCAGGCATCACAGAAGTGACAAATGACATCAGGGATCTGCCTG 594
Qy 180 GluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla 199
Db 595 GAGCAAAATAAGACCTCCAGCTGTGCGAGTTTAAGAGGACACGGGAGAGGGCTTGCC 654
Qy 200 -ArgValLeuCys---GlyGluGluGlnAlaAspAlaAspAlaGlyAla-Gln---ValC 217
Db 655 CCGAGTGTGTGCTGGGCGAGGAGCGCTGATGCTGATGCTGGGCGCCAGCGTATGC 714
Qy 217 ysSerLeuLeuAlaGlnSerGlu-ValArgPro-GlnCysLeuLeuLeuValLeuAl 236
Db 715 TCCCGGTGGTCTTGGCCAGTGTGAGCGTGAAGGCTCCAGTGTCTACTGCTGGGTCTTGC 774
Qy 236 a-----AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisG1 252
Db 775 GCCCACTAGGAACGGGAATTTGGCCCAACACGTCCTCCCAAGCTTTATCAAAAGCACCCA 834
Qy 252 nSerAspLeuLysLysLeu-----GlyIleLeuAspPheThrGluGlnAspValAlaSe 270
Db 835 ATCTGAGCCTGAACAAGCTGGGGAATCCATAGAATTTTCGCTGAGCAAGATGTTGTCGAC 894

Qy 270 rHisGlnSerTyrSerGlnLysThrLeuIleAla-----Le 282
Db 895 AAGCATAGC-----AGAAACTTATTTCCTCCCAAGAGACCCCTGATGGCACCT 945
Qy 282 uValThr---SerGlyAlaLeuLeuAlaValLeuGly-----IleThrG1 296
Db 946 GGAGACCCCTCGGGAAGCCCTGCTGGATGCTACTTGGGACTTCACGTGGGCTATTCTCC 1005
Qy 296 yTyrPheLeuMetAsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspPr 316
Db 1006 TGAGTATATATGCTC---CAGCCATCTGAGAGCCCAAGAACTGCTGGCGCAGACAA 1062
Qy 316 cTyrTyrThrGluAsnGlyGly 323
Db 1063 CCTCTTAACTCGGGATACCGGG 1084
RESULT 9
BG675028 804 bp mRNA linear EST 01-MAY-2001
LOCUS 602621258F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4746591 5',
DEFINITION mRNA sequence.
ACCESSION BG675028
VERSION BG675028.1 GI:13906424
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 804)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM10595 row: d column: 16
High quality sequence stop: 686.
FEATURES
Location/Qualifiers
source
1..804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4746591"
/clone_lib="NCI_CGAP_Skn3"
/lab_hosts="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: PCMV-SPOPT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 219 a 217 c 191 g 177 t
ORIGIN
Alignment Scores:
Pred. No.: 1.65e-75 Length: 804
Score: 994.00 Matches: 229
Percent Similarity: 91.41% Conservative: 5
Best Local Similarity: 89.45% Mismatches: 10
Query Match: 52.45% Indels: 12
DB: 10 Gaps: 2
US-09-836-602-2 (1-373) x BG675028 (1-804)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 46 ATGCCGCGGGCTGGACCGCGCTTGTCTGCTGAGTTGCTGCCTTCGGGTTCATGAGT 105
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40

```

|||||
106 CTTGACAACAACGGTACTGTCACCCAGAGTTACTACCCAGGGAACATTTCAATGTG 165
QY 41 SerThr-AsnValSerTyrGlnGlu---ThrThrThrProSerThrLeuGlySerThrSe 59
Db 166 TCTACAAAGGTATCTCTACCAAGAAAGTACAAGACCTAGGTACCCTTGTAAAGTACCAG 225
QY 59 rLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVa 79
Db 226 CTGACCCCTGTCTCAACATGGCAATGAGGCCACAACAACATCACAGAACGACAGT 285
QY 79 lLysPheThr-SerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValG 99
Db 286 CAAATTCAGCATCTACTCTGTGATAACCTCAGTTTATGGAAACACAAACTCTTCTGTCC 345
QY 99 lNserGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProG 119
Db 346 AGTCACAGACCTCTGTAATCAGACACAGTGTTCACCCAGCCCAACGTTTCAACTCCAG 405
QY 119 lUthrThr-LeuLysProSerLeuSerProGlyAsnValSerAsp-LeuSerThrThr-S 138
Db 406 AGACAGCCTTGAAGCTAGCCTGTACCTGGAAATGTTTCAGACCTTTCACACCTAG 465
QY 138 erThrSerLeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerA 158
Db 466 GCACCTAGCCTTGCACATCTCCCACTAAACCCCTATACATCATCTTCTCTATCCTAAGTG 525
QY 158 spIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleC 178
Db 526 ACATCAAGCGAAGAAATCAAAATGTCAGGCATCAGAGAAGTGAATGTGACTCAGGGCATCT 585
QY 178 ysLeuGluInAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyL 198
Db 586 GCCTGGAGCAAAATAAGACCTCCAGCTGTGCGGAGTTTAAAGAGGACAGGGGAGGGAC 645
QY 198 euAlaArgValLeuCysGlyGluGluGln-AlaAspAlaasp-AlaGlyAlaGlnVal-- 216
Db 646 TGCCCGAGTGTGTGTGGGAGGAGCAGAGCTGATGCTGATAGCTGGGGCCAGGTATA 705
QY 217 -CysSerLeuLeuAlaGlnSer-GluValArgProGlnCysLeuLeu-LeuValLeu 235
Db 706 GCTCCCTGCTACCTTGCCCGCAGCTCAGAGGTGAGGCTCAGTGTCTACTGACTGGTCCCTG 765
QY 236 -AlaAsnArgThrGluIleSerSerLys 244
Db 766 GGCAAAAGAGAAGACATTTTCAGCAAAA 793

RESULT 10
AUI35182 754 bp mRNA linear EST 24-OCT-2000
LOCUS AUI35182 PLACE1 Homo sapiens cDNA clone PLACE1001404 5', mRNA
DEFINITION AUI35182 sequence.
ACCESSION AUI35182
VERSION AUI35182.1 GI:10995721
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 754)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix

```

```

Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES             Location/Qualifiers
     source            1..754
     organism="Homo sapiens"
     db_xref="taxon:9606"
     clone="PLACE1001404"
     clone_lib="PLACE1"
     tissue_type="placenta"
     note="Vector: pME18SFL3"
BASE COUNT 201 a 210 c 172 g 166 t 5 others
ORIGIN
Alignment Scores: 2.46e-75 Length: 754
Pred. No.: 991.50 Matches: 208
Score: 991.50
Percent Similarity: 90.60%
Best Local Similarity: 88.89%
Query Match: 52.32%
Indels: 10
Gaps: 3
DB:
US-09-836-602-2 (1-373) x AUI35182 (1-754)
QY 1 MetProArgGlyThrPheAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 48 ATGCCCGGGGCTGGACCGCGCTTTGCTGCTGAGTTTGTGCTGCTTGGTTTCATGAGT 107
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 108 CTTGACAACAACGGTACTGCTACCCAGAGTTTACCTACCCAGGGAACATTTTCAAAATGTT 167
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 168 TCTACAAATGTTATCTCCACAGAAACTACACACCTAGTACCTTGGAGTACACGCTG 227
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 228 CACCCCTGTGTCTCAACATGGCAATGAGGCCACCAACAACATCACAGAAAGGACAGTCAA 287
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 288 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 347
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 348 CAGACCTCTGTANTCAGCAGAGTGTTCACCCAGCCAGCAACCTTTCACCTCCAGACA 407
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 408 ACCTTGAAGCCTAGCCTGTACCTGGAAATGTTTTCAGACCTTTCACACCTAGCAGTAGC 467
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 468 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCCTATCTTAAGTGCATCAAG 527
QY 161 Ala-GluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuG1 180
Db 528 GCAAGAAATCAATATGTTTCAGGCATCAGAGAGTGAATTCACCTACGGGCATCTGCTCGGA 587
QY 180 uGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu---GlyLeuAl 199
Db 588 GCAAAATAAGACCTCCAGCTGTGCGGAGTTTAAAGAAGGCAAGGGGAAAGGCGCTGGGC 647
QY 199 aArgVal---LeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSe 218
Db 648 CCGAGTTGCTGTGTGGGAGGAGCAAGGCTGATGCC-----TG 686
QY 218 rLeuLeuLeuAlaGlnSerGluValArgProGlnCysLeu 231
Db 687 ATTGCTGGGGGCCCAAGTTANNNGCTCCCTGNCCTCCCT 726
RESULT 11

```

CG256232	1001 bp	mRNA	linear	EST 13-FEB-2001		
LOCUS	602369136F1	NIH_MGC_91	Homo sapiens	cdna clone IMAGE:447318 5',		
DEFINITION	mRNA sequence.					
ACCESSION	CG256232					
VERSION	CG256232.1	GI:12766048				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 1001)					
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .					
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)					
JOURNAL	Unpublished (1999)					
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@hs-r@mail.nih.gov Tissue Procurement: DCTD/DTP CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAMI0305 row: p column: 23 High quality sequence start: 10 High quality sequence stop: 579.					

```

FEATURES
source
      origin      sequence      map: 575.
      Location/Qualifiers
        i. 1001
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone="IMAGE:4477318"
          /clone_lib="NIH_MGC_91"
          /tissue_type="adenocarcinoma, cell_line"
          /lab_host="DH10B (phage-resistant)"
          /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI;
          Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
          Average insert size 1.4 kb. Library enriched for
          full-length clones and constructed by Life Technologies.
          Note: this is a NIH_MGC Library."
          283 a 287 c 257 g 174 t
BASE COUNT

```

Alignment Scores:	
Pred. No.:	6,166-68
Score:	907.50
Percent Similarity:	89.87%
Best Local Similarity:	88.19%
Query Match:	47.89%
DB:	10
Length:	1001
Matches:	209
Conservative:	4
Mismatches:	15
Indels:	10
Gaps:	1

US-09-836-602-2 (1-373) x BG256232 (1-1001)

Qy 1 MetProArgGlyTrpThrAlaLeuCysYsLeuLeuSerLeuLeuProSerGlyPheMetSer 20
 |||||
Db 72 ATGCCCGGGGGCTGACCCGCCTTGCTGAGTTGCTGCCTTCTGGGTTCATGAGT 131

Qy	21	LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal	40
Db	132	CTTGACACACACGGTACTGCTACCCAGAGTTCCTACCCAGGACACATTTTCAATGTT	191

Qy	41	SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu	60
Db	192	TCTACAAATGTTATCCTACCAAGAACTACACACCTAGTACCTTGGAGTACCA	251

Qy	61	HisProvalSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys	80
Db	252	CACCCTGTGTCACATGGCAATGAGGCCAACACATCATACAGAAACGACACTCAA	311

Qy	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValcInSer 	100
Db	312	TTCACATCTACCTGTGTATAACCCTCAGTTTATGGAAACACAAAACCTTCTGTGCAGTGCA 	371

Qy	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	372	CAGACCTCTGTAATCAGCAGAGTGTCCACCAACCACCAACGTTTCAATCCAGAGACC	431
Qy	121	ThrLeuLysProSerLeuSerProGluValAsnValSerAspLeuSerThr - ThrSerThrSe	140
Db	432	ACCTTGAAGCCATGCCCTGTACCTGGAAATGTTACAGACCTTCCAAACCACTAGCACTAG	491
Qy	140	rLeuAla - ThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspile -	159
Db	492	CCTTGCAAAACATCTCCCACTAAACCCCTATACATCATCTTCTCCTATCTTAAGTGACATCA	551
Qy	160	LysAlaGluIle - LysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLe	179
Db	552	AAGGCCGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAACCTTGACTCA - GGCATCTGACT	610
Qy	179	uGluGlnAsnLysThrSerSer - Cys - AlaGluPheLysLys - AspArgGlyGluGlyL	198
Db	611	GGAGCACAAATAACACCTCCACGCTGTGTCGGAGATTTAAGAAAGGAGCAGGAGGAGGGGC	670
Qy	198	euAlaArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysS	218
Db	671	TGGCCCCAGCAGCTGTGTGGGAGGACGAGGCTGATGCGAATCGCGGGGCCCCAGGGATGCT	730
Qy	218	erLeuLeuLeuAlaGlnSerGluValArgProGln	229
Db	731	CCCCGGCTCATGACC - - - AGTCGAGGAGGAGGCCGACGG	762

RESULT	12
AU121843	
LOCUS	
DEFINITION	MAMMAL Homo sapiens CDNA clone MAMMA1001088 5', mRNA linear EST 19-OCT-2000 717 bp
ACCESSION	AU121843
VERSION	AU121843.1 GI:10937078
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 717)

TITLE
JOURNAL
COMMENT

HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Isomoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and
Yamaguchi, T.

TEL: 81-438-52-3951
Fax: 81-438-52-3952

HRI human CDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; CDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and email: genomicsnri.co.jp

```

FEATURES
source
Location/Qualifiers
1..717
/organism="Homo sapiens"
helix_research_institute:

```

```

/ob_xref=taxon:9606
/clone="MAMMAL001088"
/clone_lib="MAMMAL"
/tissue_type="mammary gland"

```

BASE COUNT	ORIGIN				
195 a	208 c	164 g	147 t		
/notes vector: pME16srL3					
					3 others

```

Alignment Scores:
Pred. No.:          3,23e-58
Score:              792.00
Matches:            163
Percent Similarity: 97.02%
Conservative:       0
Length:             717

```

Best Local Similarity: 97.02% Mismatches: 3

Query Match: 41.79% Indels: 1
 DB: 9 Gaps: 0
 US-09-836-602-2 (1-373) x AU121843 (1-717)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
 DB 214 ATGCCGGGGCTGGAGCGGCTTTGCTGCTGAGTTTGCTGCTTCTGGGTTTCATGAGT 273
 QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 DB 274 CTTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAAATGTT 333
 QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
 DB 334 TCTACAAATGATCTCCACCAAGAACTACAAACCTAGTACCCCTTGGAGTACACGCTG 393
 QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
 DB 394 CACCCCTGTGTCTCAACATGGCAATGAGGCCACAACAACATCACAGAAACGACAGTCAA 453
 QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 DB 454 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 513
 QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 DB 514 CAGACCTCTGTATACAGCAGTGTTCACCCGCCAGCCNCGTTTCAACTCCAGAGACA 573
 QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
 DB 574 ACCTTGAAGCTAGCCTGTCACTGGAAATGTTTCAGACCTTTCACACCTAGCCTAGC 633
 QY 141 LeuAlaThrSerProThrLysProThrThrSerSerSerProIleLeuSerAspIleLys 160
 DB 634 CTTTGCAACATCTTCCACTAACCCCTATCA-TCATCTCTNCTATNCTAAGTGACATCAAG 692
 QY 161 AlaGluIleLysCysSerGlyIle 168
 DB 693 GCAGAAATCAAAATGTCANGCATC 716

RESULT 13
 AL547847 997 bp mRNA linear EST 16-FEB-2001
 LOCUS AL547847 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI033YP01 5
 DEFINITION prime, mRNA sequence.

ACCESSION AL547847
 VERSION AL547847.1 GI:12882292
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 997)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1. .997

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DI033YP01"

/clone_lib="LTI_NFL006_PL2"

/tissue_type="placenta"

/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-Oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by

Life Technologies. Contact : Feng Liang Life Technologies,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com

BASE COUNT 236 a 289 c 248 g 221 t 3 others
 ORIGIN

Alignment Scores:

Pred. No.: 6,34e-58 Length: 997
 Score: 791.00 Matches: 156
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 41.74% Indels: 0
 DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x AL547847 (1-997)

QY 218 SerLeuLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsn 237
 DB 2 TCCTGCTCCTTCCGCCAGTCTGAGGTGAGGCCTCAGTGTCTACTGCTGTGTTGGCCAAC 61
 QY 238 ArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 257
 DB 62 AGAACAGAAATTTCCAGCAACTCCAACTTATGAAAAAGCACCCTGACCTGAAAAAG 121
 QY 258 LeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLys 277
 DB 122 CTGGGATCTCTAGATTCTACTGACGAGATGTTGCAAGCCACGAGAGCTATTTCCTCCAAAAG 181
 QY 278 ThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyr 297
 DB 182 ACCCTGATTGCTACTGCTCCTCGGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTAT 241
 QY 298 PheLeuMetAsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyr 317
 DB 242 TTCTCTGATGAATCGCGCAGCTGGAGCCGCCACAGAGGAAAGGCTGGCGAAGACCTTAT 301
 QY 318 TyrThrGluAsnGlyGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAla 337
 DB 302 TACACGGAACCGGTGGAGCCAGGCTATAGCTCAGGACCTGGGACCTCCCTCGAGGCT 361
 QY 338 GlnGlyLysAlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSer 357
 DB 362 CAGGGAAGGCCAGCTGTAACCCAGGGGCTCAGAAAAAGGACCGCCAGCCACCTCC 421
 QY 358 ArgAsnGlyHisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
 DB 422 AGAACGGGCATTTCAGCAAGACAAACACGTCGTGGCTGATACCGAATTG 469

RESULT 14

AU141125

LOCUS

DEFINITION

sequence.

ACCESSION AU141125

VERSION AU141125

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 733)

Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,

Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and

Isoqai, T.

HRI human cDNA project

Unpublished (2000)

Contact: Takao Isoqai

Genomics Laboratory

Helix Research Institute

1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

Tel: 81-438-52-3951

FEATURES

Source

Fax: 81-438-52-3952
 Email: genomics@hri.co.jp
 HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.

FEATURES
 source 1. .733
 Location/Qualifiers

BASE COUNT 199 a 212 c 168 g 150 t 4 others
 ORIGIN

Alignment Scores:
 Pred. No.: 4,58e-55 Length: 733
 Score: 755.50 Matches: 157
 Percent Similarity: 94.61% Conservative: 1
 Best Local Similarity: 94.01% Mismatches: 8
 Query Match: 39.87% Indels: 2
 DB: 9 Gaps: 1

US-09-836-602-2 (1-373) x AU141125 (1-733)

Qy 1 MetProArgGlyTTPThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
 Db 233 ATGCCGGGGCTGGACCGCGCTTTGCTGTGAGTTGCTGCTTCTGGGTTTCATGAGT 292
 Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 Db 293 CTTGACAAACACGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAAATGTT 352
 Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
 Db 353 TCTACAAATGTATCTACCAAGAACTACACACCTAGTACCTTGGAAAGTACCAGCGCTG 412
 Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVallys 80
 Db 413 CACCCTGTGCTCAACATGGCAATGAGCCCAACAACATCACAGAAACGACAGTCAAA 472
 Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 Db 473 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 532
 Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 Db 533 CAGACCTCTGTAATCAGCACAGTGTTCACCACCCAGCAACGTTTCAACTCCAGAGACA 592
 Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
 Db 593 ACCTTGAAGCCCTANCCGTGTCACCTGGAAATGTTTCAGACCTTTCAACCACTACACTACCT 652
 Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
 Db 653 TGCAACATCTCC---ACTAAACCTCTATCA-TCATCTTCTNCTATCTTAAAGTGACATCAAG 708
 Qy 161 AlaGluIleLysCysSerGly 167
 Db 709 GCNGAAATCAAAATGTTTCAGGC 729

RESULT 15
 LOCUS AU139201 825 bp mRNA linear EST 25-OCT-2000
 DEFINITION AU139201 PLACE1 Homo sapiens cDNA clone PLACE1010139 5', mRNA
 sequence.
 ACCESSION AU139201
 VERSION AU139201.1 GI:11000722
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 825)
 Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
 Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
 Isogai,T.

TITLE HRI human cDNA project
 JOURNAL Unpublished (2000)
 COMMENT Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp

HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.

FEATURES
 source 1. .825
 Location/Qualifiers

BASE COUNT 227 a 243 c 185 g 165 t 5 others
 ORIGIN

Alignment Scores:
 Pred. No.: 2.01e-51 Length: 825
 Score: 714.00 Matches: 148
 Percent Similarity: 95.54% Conservative: 2
 Best Local Similarity: 94.27% Mismatches: 4
 Query Match: 37.68% Indels: 3
 DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x AU139201 (1-825)

Qy 1 MetProArgGlyTTPThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
 Db 301 ATGCCGGGGCTGGACCGCGCTTTGCTGTGAGTTTGTGCTTCTGGGTTTCATGAGT 360
 Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 Db 361 CTTGACAAACACGGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAAATGTT 420
 Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
 Db 421 TCTACAAATGTATCTACCAAGAACTACACACCTAGTACCTTGGAAAGTACCAGCGCTG 480
 Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVallys 80
 Db 481 CACCCTGTGCTCAACATGGCAATGAGGCCACAAACAACATCACAGAAACGACAGTCAAA 540
 Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 Db 541 TTCACATCTACTCTGTGATAACCTCAGTTTATGGAACAAACATCTTGTCTCCAGTCA 600
 Qy 101 Gln-ThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 Db 601 CAAGACCTCTGTAATCAGCACAGTGTTCANCAACCCACCAACGTTTCAACTCCAGAGAC 660
 Qy 120 rThrLeuLysProSerLeuSerProGly-AsnValSerAspLeuSerThrThrSerThrS 140
 Db 661 AACCTTGAAGCCTAACCTGTACCTGGAAATGTTTCAACGCTTTTAACCACTTACACTA 720
 Qy 140 rLeuAlaThrSerProThr-LysProTyrThrSerSerSerPro 154
 Db 721 GCCTTGCACATCTCCCACTAAACCCCTATACATCATCTCTCTCT 765

Search completed: October 30, 2002, 10:47:52
Job time : 1301.57 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 30, 2002, 08:08:13 : Search time 47 8495 Seconds
(without alignments)
1914.781 Million cell updates/sec

Title: US-09-836-602-2
Perfect score: 1895
Sequence: 1 MPRGWTALCLSLPSGFMS.....QATSRNGHSARQHVADTEL 373

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 383533 seqs, 122816752 residues
Total number of hits satisfying chosen parameters: 767066
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09836602/runat.29102002_101145_3837/app_query.fasta.1.1429
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptlo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602 -CGN1_1_44 -runat.29102002_101145_3837 -NCFU=6 -ICPU=3
-NO_XLPAY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6 -FCAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents.NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6C.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	904	47.7	1065	2	US-08-475-634D-18
2	144	7.6	5163	3	US-08-700-651-1
3	144	7.6	5163	3	US-08-928-361B-4
4	144	7.6	5318	3	US-08-700-651-2
5	144	7.6	5318	3	US-08-928-361B-3
6	138	7.3	5511	3	US-08-928-361B-2
7	138	7.3	7334	3	US-08-928-361B-1
8	129	6.8	1505	1	US-07-915-246-1
9	128	6.8	390	4	US-09-197-649-7
10	127.5	6.7	3337	1	US-08-072-610-1
11	127.5	6.7	3337	2	US-08-719-822B-1
12	127.5	6.7	3337	4	US-09-092-458-1

13	125.5	6.6	2589	1	US-08-325-267A-3	Sequence 3, Appli
14	125.5	6.6	2685	3	US-08-362-525-21	Sequence 21, Appl
15	125.5	6.6	4614	1	US-08-325-267A-1	Sequence 1, Appli
16	123	6.5	2754	1	US-08-270-076A-10	Sequence 10, Appl
17	121	6.4	2214	6	5258502-1	Patent No. 5258502
18	121	6.4	3168	4	US-09-165-239A-3	Sequence 3, Appli
19	119.5	6.3	688	4	US-08-998-416-915	Sequence 915, App
20	119.5	6.3	1107	2	US-08-991-300-1	Sequence 1, Appli
c 21	119.5	6.3	80161	3	US-09-036-987A-1	Sequence 1, Appli
c 22	119.5	6.3	80161	4	US-09-370-700-1	Sequence 1, Appli
23	118	6.2	2150	2	US-08-861-464-13	Sequence 13, Appl
24	118	6.2	2150	2	US-08-396-001-13	Sequence 13, Appl
25	118	6.2	2150	4	US-09-323-433A-13	Sequence 13, Appl
26	117	6.2	2584	3	US-08-758-662-8	Sequence 8, Appli
c 27	116	6.1	2793	1	US-08-209-747-1	Sequence 1, Appli
c 28	116	6.1	2793	1	US-08-458-298-1	Sequence 1, Appli
29	115.5	6.1	2032	4	US-09-241-581B-5	Sequence 5, Appli
30	115.5	6.1	2032	5	PCT-US95-07721-5	Sequence 5, Appli
31	114.5	6.0	2093	1	US-08-287-001A-1	Sequence 1, Appli
32	114.5	6.0	2093	5	PCT-US95-09941-1	Sequence 1, Appli
33	113.5	6.0	2164	4	US-08-760-615-3	Sequence 3, Appli
34	113	6.0	36519	3	US-08-923-137-2	Sequence 2, Appli
35	112.5	5.9	4031	2	US-08-993-118-1	Sequence 1, Appli
36	112.5	5.9	4031	3	US-08-845-528C-1	Sequence 1, Appli
37	112.5	5.9	4225	2	US-08-993-118-9	Sequence 9, Appli
38	112.5	5.9	4225	3	US-08-845-528C-9	Sequence 9, Appli
39	112.5	5.9	4265	4	US-09-061-709-1	Sequence 1, Appli
c 40	111	5.9	3181	1	US-08-655-086-1	Sequence 1, Appli
c 41	110	5.8	6027	2	US-08-968-542C-1	Sequence 1, Appli
42	109.5	5.8	3833	1	US-08-917-320-18	Sequence 18, Appl
43	109.5	5.8	3833	5	PCT-US95-04611A-18	Sequence 18, Appl
44	109.5	5.8	5931	3	US-08-783-774-1	Sequence 1, Appli
45	108.5	5.7	4108	4	US-08-981-729-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-475-634D-18
; Sequence 18, Application US/08475634D
; Patent NO. 5962644
; GENERAL INFORMATION:
; APPLICANT: HAWLEY, Robert J.
; APPLICANT: MONROY, Rodney L.
; TITLE OF INVENTION: Antibodies to Porcine CD34 Positive
; TITLE OF INVENTION: and Uses Therefor
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi, Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS 6.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475.634D
; FILING DATE: June 7, 1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24, 025
; REFERENCE/DOCKET NUMBER: 61750-141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 973-994-1700
; TELEFAX: 973-994-1744
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:


```
Qy 184 ThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArgValLeuCys 203
Db 3295 ---GGCGAATGT-----GAGCAGAAAGGAGCAACTTATGTTGGTGT 3333
Qy 204 GlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeuLeuAlaGln 223
Db 3334 ATCGGAAAGATGACGACGATTTGAAATGGAATGTCATTTACAATGATTTCCAAATGATGAC 3393
Qy 224 SerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGluIle----- 241
Db 3394 ACGCATGTCCGTTTCAGATTTAAGTAAAGATGATGAGACACTATTTCAGTAAAGATGC 3453
Qy 242 -----SerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeu 258
Db 3454 GGAAGAGTGCAGGTAAACTCGAGTTCACAGATAGAAGT----- 3492
Qy 259 GlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTySerGlnLysThr 278
Db 3493 -----TTGGATTTTCAACAATTCCTCCAGTAGCTGGCCATAACAGCTGTTCAC----- 3537
Qy 279 LeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyPhe 298
Db 3538 -----ATATAGTTGGTGTGAGCGGC----- 3558
Qy 299 LeuMetAsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTy 317
Db 3559 -----GGTGGAAAAATTCACGTAAGCCCATAC 3585

RESULT 8
US-07-915-246-1/c
; Sequence 1, Application US/07915246
; Patent No. 5401836
; GENERAL INFORMATION:
; APPLICANT: Baszczynski, Chris L.
; APPLICANT: Fallis, Lynne
; APPLICANT: Bellmare, Guy
; APPLICANT: Boivin, Rodolphe
; TITLE OF INVENTION: A BRASSICA REGULATORY SEQUENCE FOR
; TITLE OF INVENTION: ROOT-SPECIFIC OR ROOT-ABUNDANT GENE EXPRESSION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McAndrews, Held, and Malloy
; STREET: 500 W. Madison St. 34th Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/915,246
; FILING DATE: 19920716
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 91 P 1125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 707-8889
; TELEFAX: 312 707-9155
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1505 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEtical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
```

```
; ORGANISM: Brassica napus
; STRAIN: Westar
; DEVELOPMENTAL STAGE: Somatic
; TISSUE TYPE: Root
; US-07-915-246-1

Alignment Scores:
Pred. No.: 0.00104 Length: 1505
Score: 129.00 Matches: 47
Percent Similarity: 48.78% Conservatives: 33
Best Local Similarity: 28.66% Mismatches: 70
Query Match: 6.81% Indels: 14
DB: 1 Gaps: 5

US-09-836-602-2 (1-373) x US-07-915-246-1 (1-1505)

Qy 12 SerLeuLeuProSerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeu 31
Db 1370 AGCTGCATAACCTCCCTCCTCCTCCGCCACCTCCACGCTCCACCCACGCGCTCCTCC 1311
Qy 32 ProThrGlnGlyThrPheSer-----AsnValSerThrAsnValSerTyr 46
Db 1310 ACCACCTGTCTCCTCTCTCCACCCACCCACCGTATCCCTCTCCAGCTCCACCTCCGGC 1251
Qy 47 GlnGluThrThrProSerThrLeuGlySerThrSerLeuHisProValSerGlnHis 66
Db 1250 ACCATAACCACCGGTGAGCTCCACCGCGCGCAGAACCTCTCTCCACGCGCTCTTT 1191
Qy 67 GlyAsnGluAlaThrThrAsnIle-----ThrGluThrThrValLysPheThrSer 83
Db 1190 TCCCGCGCTCCACCGCTCCATGTCCACCTGCACCGCCACCTCCATATCTCTCCAGC 1131
Qy 84 ThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSerGlnThrSer 103
Db 1130 ACCAGCGCTTCTCCACCGGTATCCACCGCTGTCACCTCCACCTCCACCTCCACCGCC 1071
Qy 104 ValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThrThrLeuLys 123
Db 1070 TCCATTTCCACCGCTCCACCGC-----TCCATGTCACCTGCACCGCCACCC----- 1023
Qy 124 ProSerLeuSerProGlyAsnValSerAspLeuSerThrThr-----SerThrSerLeuAla 142
Db 1022 ---TCCATATCTCTCCTCCAGCAGCAGCAGCTTCTCCACCGCCACCTCCACCGCTCCACCG 966
Qy 143 ThrSerProThrLysProTyThrSerSerSerProIleLeuSerAspIleLysAlaGlu 162
Db 965 ACCGCGAGCTCTCCACCGCAGCAGCTCTCTCCACCGCCACCTCTCCACCTCCACCATGCTCC 906
Qy 163 IleLysCysSer 166
Db 905 TCCCTCCGCTCC 894

RESULT 9
US-09-197-649-7
; Sequence 7, Application US/09197649
; Patent No. 6194550
; GENERAL INFORMATION:
; APPLICANT: Gold, Larry
; APPLICANT: Tuerk, Craig
; APPLICANT: Pribnow, David
; APPLICANT: Smith, Jonathan D.
; TITLE OF INVENTION: Systematic Polypeptide Evolution by Reverse Translation
; FILE REFERENCE: NEX02/C1-CON
; CURRENT APPLICATION NUMBER: US/09/197,649
; CURRENT FILING DATE: 1998-11-23
; EARLIER APPLICATION NUMBER: 07/829,461
; EARLIER FILING DATE: 1992-01-31
; EARLIER APPLICATION NUMBER: 07/739,055
; EARLIER FILING DATE: 1991-08-01
; EARLIER APPLICATION NUMBER: 07/561,968
; EARLIER FILING DATE: 1990-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
```


Qy 294 IlerThrGlyTyrPheLeuMetAsnArgArgSerTrpSerPro-----ThrGlyGluArg 311
Db 1963 GTTACTACAGATAC-----ACACATGGTGTCTCTATTTCACACAGAAATCG 2010
Qy 312 LeuGlyGluAspProTyrTyrThrGluAsnGlyGlyGlnGlyTyrSerSerGly--- 330
Db 2011 AGGCAACAACAAACGCTAGTTACTTCTCTCGAATCTGGTGTGTCCGAAACT 2070
Qy 330 ----- 330
Db 2071 GCTTCACTGCCATGTTTCGAGCCACGCTACTGTGAATGATGTTGTTACGGTCTAT 2130
Qy 331 ProGlyThrSerProGluAlaGlnGlyLysAlaSerValAsnArgGlyAlaGlnGluAsn 350
Db 2131 CCTACATGGAGCCACAGACTCGAATGAAGAGTCTGTGAGC---TCTAAATGAACAGT 2187
Qy 351 GlyThrGlyGlnAlaThrSerArgAsn-----GlyHisSerAlaArgGlnHisValVal 368
Db 2188 GCTACCGGTGAGACAAACCACTATCTTAGCTGTGTAACGACTACCAATACTGTAGCT 2247
Qy 369 AlaAspThr 371
Db 2248 GCTGAGACG 2256

RESULT 14
US-08-362-525-21
; Sequence 21, Application US/08362525
; Patent No. 6027910
; GENERAL INFORMATION:
; APPLICANT: KLIS, FRANCISCUS M.
; APPLICANT: SCHREUDER, MAARTEN P.
; APPLICANT: TOSCHKA, HOLSER Y.
; APPLICANT: VERRIPS, CORNELIS T.
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,525
; FILING DATE: 04-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92202080.5
; FILING DATE: 08-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92203899.7
; FILING DATE: 14-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/01763
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 213289/T7020(V)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 2685 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; IMMEDIATE SOURCE:
; CLONE: pY105
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2685
; OTHER INFORMATION: /product= "Flocculation protein" /gene= nFLOI"
US-08-362-525-21
Alignment Scores:
Pred. No.: 0.00655 Length: 2685
Score: 125.50 Matches: 100
Percent Similarity: 38.30% Conservative: 62
Best Local Similarity: 23.64% Mismatches: 148
Query Match: 6.62% Indels: 115
DB: 3 Gaps: 20

US-09-836-602-2 (1-373) x US-08-362-525-21 (1-2685)
Qy 26 ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnValSer 45
Db 1234 ACAACCACCACTGAA---CCATGGAGCTGGTACTTTTACTTCGACTTCACCTCAAAATGCT 1290
Qy 46 -----TyrGlnGluThrThr-----ThrPro 52
Db 1291 ACTGTCACCTGGAACCAATGGCTTGCCAACTGATGAACCTGATGTTGTTCAAACTCCA 1350
Qy 53 SerThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThr 72
Db 1351 ACTACTGCATCTCATCCAGTTGTGCA-----TCATCATCTTCAGGACAAATCACCAGC 1404
Qy 73 AsnIleThrGluThrThrValLysPheThrSerThrSerValIleThrSerValThrGly 92
Db 1405 TCTATCAGC-----TCTTCGCGTCCCAATATTATCCCATCTCTATCTCT 1446
Qy 93 AsnThrAsnSerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrPro 112
Db 1447 AGCAATGGAACCTCTGTGATTCT---TCCCTCAGTAATTTCT----- 1485
Qy 113 AlaAsnValSerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSer 132
Db 1486 TCCTCAGTCACTTCTCTTATTCATCTTCTTCCAGTCAATTTCTCTCCTCAGTCAATTTCT 1545
Qy 133 AspLeuSerThrThrSerThrSerLeu-----AlaThrSerPro 145
Db 1546 TCTTCTACAACAACTCCACTCTATATTTCTGAATCATCTAAATCATCCGTCATTTCCA 1605
Qy 146 ThrLysProTyrThrSerSerSerProIleLeuSerAspIleLysAlaGluLysCys 165
Db 1606 ACCAGTAGTTCACCTCTGGTTCT-----TCTGAGAGCGAAAGC 1644
Qy 166 SerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGluGlnAsnLysThrSer 185
Db 1645 AGTTCAGCTGGTCT 1704
Qy 186 -----SerCysAlaGluPheLysLysAspArgGly 195
Db 1705 ACATATTTCTTCTTCATCATTTACCATTTGTTACCGGACCAACAGCGAGAAACTGCT 1764
Qy 196 GluGlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAla----- 210
Db 1765 TCTTCAATAC---CAGCTGCTACCACTACAAAAACGAGCGAACAACCACTTTGGTTACCG 1822
Qy 211 ---AspAlaGlyAlaGlnValCysSerLeuLeuLeuLeuGlnSerGluValArgProGln 229
Db 1823 TGACATCTCGGAGTCTCATGTGTGCAGTGAATTCATCTCCCTCGGATTTGTTTCCACAG 1882

```
QY 230 CysLeuLeuValLeuAlaAsnArgThrGluile-----SerSerLysLeuGlnLeuMet 248
|||||
Db 1883 CTACTGTTACTGTTAGCGGCTCACACAGAGTATACCACATGGTGCCTATTCTACTA 1942
|||||
QY 249 LysLysHisGlnSerAspLeuLysLeuGlyLeuAspPheThr----- 264
|||||
Db 1943 CAGAGACAACAAGC-----AAACCAAA-GGGACAACAGCAACCAACAGAAACAACA 1995
|||||
QY 265 -----GluGlnAspValAlaSerHisGlnSer 273
|||||
Db 1996 AAACAAACCCAGGAGTACAAATTTCTTGTGGAATCTGACGTATGCTCT----- 2046
|||||
QY 274 TyrSerGlnLysThrLeuAlaLeuValThrSerGlyAlaLeuValLeuGly 293
|||||
Db 2047 -----AAGCTGCTTCTCCAGCCATTGTATCTACAGCCTGCTACTATTAAACGGC 2097
|||||
QY 294 IleThrGlyTyrPheLeuMetAsnArgSerTrpSerPro-----ThrGlyGluArg 311
|||||
Db 2098 GTTACTACAGAAATAC-----ACAACATGGTGTCTCTATTTCACACACAGAAATCG 2145
|||||
QY 312 LeuGlyGluAspProTyrTyrThrGluAsnGlyGlyGlnGlyTyrSerSerGly--- 330
|||||
Db 2146 AGGCACAACAACAGCGTAGTTACTGTTACTTCTCGGAATCTGGTGTTCGAAACT 2205
|||||
QY 330 ----- 330
|||||
Db 2206 GCTTCACCTGCCATTGTTTCGACGCCACGGCTACTGTGAATGATGTTGTACGGTCTAT 2265
|||||
QY 331 ProGlyThrSerProGluAlaGlnGlyLysAlaSerValAsnArgGlyAlaGlnGluAsn 350
|||||
Db 2266 CCTACATGGAGGCCACACAGCTCGGAATGAAGAGTCTGTCTCAGC---TCTAAATGAACAGT 2322
|||||
QY 351 GlyThrGlyGlnAlaThrSerArgAsn-----GlyHisSerAlaArgGlnHisValVal 368
|||||
Db 2323 GCTACCCGGTGGAGACACACCAATACTTTAGCTGCTCGAAGCAGCTACCAATCTGTAGCT 2382
|||||
QY 369 AlaAspThr 371
|||||
Db 2383 GCTGAGACG 2391
|||||
RESULT 15
US-08-325-267A-1
; Sequence 1, Application US/08325267A
; Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI
; APPLICANT: TAKATA, YOSHIHIRO
; APPLICANT: OGAWA, MASAHIRO
; APPLICANT: PENTTILA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, SIRKKA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; TITLE OF INVENTION: CONTAINING THEM
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,267A
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP PCT/JP94/00290
; FILING DATE: 24-FEB-1994
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 38871/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4614 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORGANISM: Saccharomyces cerevisiae
; STRAIN: ABXL-1D
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4614
; US-08-325-267A-1
Alignment Scores:
Pred. No.: 0.0168 Length: 4614
Score: 125.50 Matches: 100
Percent Similarity: 38.30% Conservative: 62
Best Local Similarity: 23.64% Mismatches: 148
Query Match: 6.62% Indels: 115
DB: 1 Gaps: 20
US-09-836-602-2 (1-373) x US-08-325-267A-1 (1-4614)
QY 26 ThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnValSer 45
|||||
Db 3124 ACAACCAACCACTGAA---CCATGGACTGGTACTTTTACTTCGACTTCCACTGAAATGTCT 3180
|||||
QY 46 -----TyrGlnGluThrThr-----ThrPro 52
|||||
Db 3181 ACTGTCTACTGGAACCAATGCTTGCACCACTGATGAAACTGTCATTTGTGTCACCAACTCCA 3240
|||||
QY 53 SerThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThr 72
|||||
Db 3241 ACTACTGCCATCTCATCCAGTTTGTCA-----TCATCATCTTCAGGACAAATCACCAGC 3294
|||||
QY 73 AsnIleThrGluThrThrValLysPheThrSerThrSerValIleThrSerValTyrGly 92
|||||
Db 3295 TCTATCAGC-----TCTTCGCGTCCAATTTATTACCCCATTTCTATCTCT 3336
|||||
QY 93 AsnThrAsnSerSerValGlnSerGlnThrSerValIleSerThrValPheThrThrPro 112
|||||
Db 3337 AGCAATGGAACTTCTGTGATTCT---TCCTCAGTAATTTCT----- 3375
|||||
QY 113 AlaAsnValSerThrProGluThrThrLeuLysProSerLeuSerProGlyAsnValSer 132
|||||
Db 3376 TCTCAGTCACTTCTCTCTATTCACACTTCTCTCCAGTCAATTTCTCTCAGTCATTTCT 3435
|||||
QY 133 AspLeuSerThrThrSerThrSerLeu-----AlaThrSerPro 145
|||||
Db 3436 TCTTCTACAACCACTCCACTTCTATATTTTCTGAATCATCAATCAATCCGTCATTTCCA 3495
|||||
QY 146 ThrLysProTyrThrSerSerSerProIleLeuSerAspIleLysAlaGluIleLysCys 165
|||||
Db 3496 ACCAGTAGTTCCACCTCTGTGTTCT-----TCTGAGACGCAACG 3534
|||||
QY 166 SerGlyIleArgGluGluLysLeuThrGlnGlyIleCysLeuGluGlnAsnLysThrSer 185
|||||
Db 3535 AGTTCAGCTGGTCTGTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3594
|||||
QY 186 -----SerCysAlaGluPheLysLysAspArgGly 195
|||||
```

Search completed: October 30, 2002, 10:50:26
Job time : 68.8495 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:04:33 ; Search time 169,922 Seconds
(without alignments)
3768.837 Million cell updates/sec

Title: US-09-836-602-2

Perfect score:

Sequence: 1 MPRGWTALCLSLPSGEMS.....QATSRNGHSARQHVADTEL 373

Scoring table: BLOSUM62

Xgapop	10.0	Xgapext	0.5
Ygapop	10.0	Ygapext	0.5
Fgapop	6.0	Fgapext	7.0
Delop	6.0	Delext	7.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

Command line parameters:

```
-MODEL=frames_p2n_model -DEV=xlp
-Q/cgcn2_1/USPTO.spool/US09836602/runat_29102002.101143_3758/app_query_fasta_1.1429
-DB=NCGenesec.032802 -QFMT=FASTAP -SUFFIX=ring -MINMATCH=0.1 -LOOPL=0
-LOOPT=0 -UNITS=bits -START=1 -END=-1 -WATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptn -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602_@CGN_1_312_@runat_29102002.101143_3758 -NCPU=6 -ICPU=3
-NO_XLPX -NO_MMAP -LARGEOQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

```

Database :
1: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1980.DAT:**
2: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1981.DAT:**
3: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1982.DAT:**
4: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1983.DAT:**
5: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1984.DAT:**
6: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1985.DAT:**
7: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1986.DAT:**
8: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1987.DAT:**
9: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1988.DAT:**
10: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1989.DAT:**
11: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1990.DAT:**
12: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1991.DAT:**
13: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1992.DAT:**
14: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1993.DAT:**
15: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1994.DAT:**
16: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1995.DAT:**
17: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1996.DAT:**
18: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1997.DAT:**
19: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1998.DAT:**
20: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA1999.DAT:**
21: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA2000.DAT:**
22: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:**
23: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA2001B.DAT:**
24: /SIDS1/cgcdata/geneseq/geneseqn-emb1/NA2002.DAT:**

```

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1895	100.0	1122	24	Human full-length
2	1895	100.0	2463	22	Nucleotide sequenc
3	1891	99.8	2615	21	Human CD-34 polynu
4	1891	99.8	2615	21	Human adenosine re
5	1891	99.8	3490	21	Human CD-34 polynu
6	1891	99.8	3490	21	Human adenosine re
7	1574	83.1	951	24	Human CD34 surface
8	1509	79.6	906	24	Human CD34 surface
9	904	47.7	1065	18	Porcine CD34 from
10	508	26.8	615	21	Human secreted pro
11	311	16.4	233	22	Human breast cell
12	311	16.4	233	22	Human foetal liver
13	311	16.4	233	22	Probe #14530 for g
14	311	16.4	233	22	Human brain expres
15	311	16.4	233	22	Human bone marrow
16	311	16.4	233	22	Probe #13951 for g
17	311	16.4	233	22	Probe #18003 used
18	311	16.4	233	22	Probe #9593 used t
19	265	14.0	209	22	Human breast cell
20	265	14.0	209	22	Human foetal liver
21	265	14.0	209	22	Probe #14140 for g
22	265	14.0	209	22	Human brain expres
23	265	14.0	209	22	Human bone marrow
24	265	14.0	209	22	Probe #13530 for g
25	265	14.0	209	22	Probe #17596 used
26	265	14.0	209	22	Probe #9203 used t
27	233	12.3	235	22	Novel human diagno
28	174	9.2	406	22	Human breast cell
29	174	9.2	406	22	Human foetal liver
30	174	9.2	406	22	Probe #4241 for ge
31	174	9.2	406	22	Human brain expres
32	174	9.2	406	22	Human bone marrow
33	174	9.2	406	22	Probe #4329 for ge
34	174	9.2	406	22	Probe #4456 used t
35	174	9.2	406	22	Probe #4210 used t
36	159	8.4	13154	20	Enterococcus faeca
37	152	8.0	2131	22	Murine PCPL1 codin
38	150.5	7.9	2336	23	Drosophila melanog
39	148.5	7.8	5848	24	Human DNA sequenc
40	148	7.8	40875	18	Insert from cosmid
41	147.5	7.8	7720	21	Genomic DNA encodi
42	146	7.7	4116	23	Drosophila melanog
43	146	7.7	5894	22	Human kidney cell
44	146	7.7	6116	23	Drosophila melanog
45	146	7.7	7035	23	Staphylococcus aur

ALIGNMENTS

RESULT 1

RESUL I
AAD22661

AAD22001
IP AAD22661 standard: DNA: 1122 bp.

XX
10077200

AC AAD22661;

XX
/ TOC/

DT 26-FEB-2002 (first entry)

XX
XX
XX

DE Human full-length CD34 (flCD34) surface antigen encoding DNA.

XX

KW Human; surface marker; surface

KW full-length CD34; f1CD34; ds.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qual
----	-----	---------------

FT XX /product= "Human full-length CD34 surface antigen"
 PN EPI148066-A1.
 XX 24-OCT-2001.
 PD 18-APR-2001; 2001EP-0109374.
 PF 18-APR-2000; 2000DE-1019075.
 PR (ZAND/) ZANDER A R.
 XX zander AR;
 PI WPI: 2002-019289/03.
 DR P-PSDB; AAE13541.
 XX
 PT New gene transfer vector (accession number DSM13396) containing a
 PT transgene and a nucleic acid sequence coding for a surface marker,
 PT useful in gene therapy, and for detecting genetically modified cells or
 PT cells which do not express CD34 -
 XX
 PS Claim 3; Page 9-11; 28pp; English.
 XX
 CC The patent discloses a gene transfer vector (accession number DSM13396)
 CC containing a transgene and a nucleic acid sequence coding for a surface
 CC marker. The surface marker is the CD34 surface antigen, its fragment or
 CC variant. The vector is useful for in vitro transduction of T lymphocytes,
 CC for gene therapy, and in the enrichment, detection and analysis of cells
 CC in vitro that do not naturally express CD34. T lymphocytes transduced
 CC with the vector are also useful in gene therapy. The CD34 nucleic acid
 CC sequences (marker genes), their fragments or variants are used for
 CC detecting genetically modified cells or cells which do not naturally
 CC express CD34. The present sequence is a DNA encoding human full-length
 CC CD34 (fLCD34) surface antigen.
 XX
 SQ Sequence 1122 BP; 305 A; 311 C; 270 G; 236 T; 0 other;

 Alignment Scores:
 Pred. No.: 6,89e-138 Length: 1122
 Score: 1895.00 Matches: 373
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 24 Gaps: 0

 US-09-836-602-2 (1-373) x AND22661 (1-1122)

 Qy 1 MetProArgGlyThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
 DB 1 ATGCCGGGGCTGCACCGCGCTTTGCTGTGAGTTTGCTGCTTCTGGGTTTCATCAGT 60

 Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 DB 61 CTGTGACAAACACGGTACTGTACCCAGAGTTTACCTACCCAGGGAACATTTTCAAAATGTT 120

 Qy 41 SerThrAsnValSerTyrglnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
 DB 121 TCTACAANTGTATCTACCAAGAACTACAAACCTAGTACCTTGGAGTACCAGCGCTG 180

 Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrVallys 80
 DB 181 CACCCTGTCTCAACATGCAATGCAATGAGGCCACAAACATCACAGAAACGACAGTCAAA 240

 Qy 81 PheThrSerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValGlnSer 100
 DB 241 TTCACATCTACCTCTGTGATAACCTCAGTTATGGAACACAAACTCTTCTGTCCAGTCA 300

 Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 DB 301 CAGACCTCTGTATACACAGAGTGTTCACCACCCGCCCAACGTTTCAACTCCAGAGACA 360

 Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140

Db 361 ACCTTGAAGCCTAGCCTGTACCTGGAAATGTTTACAGACCTTTCAACCACTAGCACTAGC 420
 Qy 141 LeuAlaThrSerProThrLysProTyThrSerSerProIleLeuSerAspIleLys 160
 Db 421 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTCTATCTCCCTAAAGTGACATCAAG 480
 Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
 Db 481 GCAGAAATCAATGTTTCAGGCATCAGAGAAAGTGAATTTGACTCAGGGCATCTGCCTGGAG 540
 Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
 Db 541 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAAGCACAGGGAGAGGCGCTGGCCCGA 600
 Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
 Db 601 GTGCTGTGTGGGAGGAGCAGGTGATGCTGCTGGGGCCCAAGGTATGCTCCCTGCTC 660
 Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
 Db 661 CTTGCCCACTGTAGGTGAGGCCCTCAGTGCTACTGCTGTGCTTGGCCCAACAGACAGAA 720
 Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
 Db 721 ATTTCCAGCAAACTCCAACTTATGAAAAGACCACTGACCTGAAAAAGTGGGATC 780
 Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrrSerGlnLysThrLeuIle 280
 Db 781 CTAGATTTCACTGAGCAAGATGTGCAAGCCACAGAGCTATTCCCAAAAGACCCCTGATT 840
 Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrrPheLeuMet 300
 Db 841 GCAGTGTACCTCGGAGGCCCTGCTGCTGTGCTTGGGCATCACTGGCTATTTCTCTGATG 900
 Qy 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyrrTyThrGlu 320
 Db 901 AATCGCGCAGCTGGAGCCCAAGAGGCTGGCGGAAAGGCTGGCGAAGACCCCTATTACACGGA 960
 Qy 321 AsnGlyGlyGlnGlyTyrrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys 340
 Db 961 AACGGTGGAGCCAGGCTATAGCTCAGGACCTGGGACCTCCCTCAGGCTCAGGGAAAG 1020
 Qy 341 AlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGly 360
 Db 1021 GCCAGTGTGAACCGAGGGCTCAGGAAACGGGACCGCCAGGCCACCTCCAGAAACGGC 1080
 Qy 361 HisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
 Db 1081 CATTGACAAAGACACACGTTGGTGGCTGATACCGAATTG 1119

 RESULT 2
 AAH75132 standard; DNA; 2463 BP.
 XX
 XX AAH75132;
 XX
 XX 13-NOV-2001 (first entry)
 XX
 XX Nucleotide sequence of a human CD34 polypeptide.
 DE
 XX Human; CD34 gene; blast.crisis; chronic myelogenous leukemia;
 KW nm23-H4 kinase gene; ss.
 XX
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FT CDS 91..1248
 FT /*tag= a
 FT /product= "CD34"
 XX
 PN WO200164946-A1.
 XX

PA (UYEC-) UNIV EAST CAROLINA.
 PA (NYCE/) NYCE J W.
 XX Nyce JW;
 XX WPI; 2000-679539/66.
 XX
 XX Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -
 XX
 PS Disclosure; Page 292-293; 1592pp; English.
 XX

CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS), and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impeded respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.
 XX

SQ Sequence 2615 BP; 616 A; 763 C; 627 G; 609 T; 0 other;

Alignment Scores:
 Pred. No.: 4.32e-137 Length: 2615
 Score: 1891.00 Matches: 372
 Percent Similarity: 100.00% Conservatives: 1
 Best Local Similarity: 99.73% Mismatches: 0
 Query Match: 99.79% Indels: 0
 DB: 21 Gaps: 0

US-09-836-602-2 (1-373) x AAF20898 (1-2615)

Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
 |||||
 Db 294 ATGCCGCGGGCTGGACCGCGCTTTGCTGCTGAGTTTGCTGCTTCTGGGTTTCATGAT 353
 |||||
 Qy 21 LeuAspAsnGlyThrAlaThrProGluLeuProGluThrGlnGlyThrPheSerAsnVal 40
 |||||
 Db 354 CTTGACAAACACGGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAAATGTT 413
 |||||
 Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
 |||||
 Db 414 TCTACAATGTATCTCTACCAAGAAACTACACACCTAGTACCCCTGGAGTACCAGCGCTG 473
 |||||
 Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
 |||||
 Db 474 CACCCCTGTGCTCAACATGCAATGAGGCCACACAAACATCACAGAAGACACAGTCAAA 533
 |||||
 Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 |||||

Db 534 TTCACATCTACCTCTGTGATAAACCTCAGTTTATGGAAACACAAACTCTTCTGTCCAGTCA 593
 Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 |||||
 Db 594 CAGACCTCTGTGTAATCAGACAGAGTGTTCACCACCCACCAACGTTTCAACTCCAGAGACA 653
 |||||
 Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
 |||||
 Db 654 ACCTTGAAGCTACGCTGTACCTGGAAATGTTTCAGACCTTTCACACCTAGCAGCTAGC 713
 |||||
 Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
 |||||
 Db 714 CTTGCAACATCTCCACCTAAACCTTATACATCATCTCTCTCTATCTTCCTAAGTGACATCAAG 773
 |||||
 Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
 |||||
 Db 774 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAAATTTGACTCAGGGCATCTGCTCGGAG 833
 |||||
 Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
 |||||
 Db 834 CAAATAAGACCTCCACCTGTGCGGAGCTTTAAGAAGACACAGGGGAGAGGCGCTGGCCCGA 893
 |||||
 Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
 |||||
 Db 894 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGCTGGGCCAGGATGCTCCCTGCTC 953
 |||||
 Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
 |||||
 Db 954 CTTGCCAGCTGTAGGTGAGGCTCAGTGCTACTGTGCTGTGGCCACAGAACAGAA 1013
 |||||
 Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
 |||||
 Db 1014 ATTTCCAGCAACTCCAACTTATGAAAGACCACTGACCTGACCTGAAAGAGCTGGGATC 1073
 |||||
 Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
 |||||
 Db 1074 CTAGATTTCTAGTACGCAAGATGTTGCAAGCCACACAGAGCTATTTCCCAAGACCTGATT 1133
 |||||
 Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
 |||||
 Db 1134 GCACCTGCTACCTCGGAGGCGCTGCTGCTGCTGCTGGGCATCACTGGCTATTTCTCATG 1193
 |||||
 Qy 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyrTyrThrGlu 320
 |||||
 Db 1194 AATCGCGCAGCTGGAGCCGCACAGAGAAAGGCTGGCGGAAGACCTTATTACACGGAA 1253
 |||||
 Qy 321 AsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys 340
 |||||
 Db 1254 AACGGTGGAGGCCAGGCGCTATAGCTCAGGACCTGGGACCTCCCTCAGGCTCAGGGGAAAG 1313
 |||||
 Qy 341 AlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGly 360
 |||||
 Db 1314 GCCAGTGTGAACCGAGGGGCTCAGAAAACGGGACCGCCAGGCGCACCTCCAGAAACGGC 1373
 |||||
 Qy 361 HisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
 |||||
 Db 1374 CATTCAGCAAGACAAACACGTGGTGGCTGATACCGAATTG 1412
 |||||
 RESULT 4
 AAA34776
 ID AAA34776 standard; DNA; 2615 BP.
 XX
 AC AAA34776;
 XX
 XX 28-JUL-2000 (first entry)
 DT
 XX Human adenosine receptor related polynucleotide SEQ ID NO:2465.
 DE
 XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytostatic; analgasic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;

respiratory distress syndrome; pain; cystic fibrosis; emphysema; pulmonary hypertension; chronic obstructive pulmonary disease; COPD; cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

Homo sapiens.

WO200009525-A2.

24-FEB-2000.

03-AUG-1999; 99WO-US17712.

03-AUG-1998; 98US-0095212.

(UYEC-) UNIV EAST CAROLINA.

Nyce JW;

WPI; 2000-205971/18.

New antisense oligonucleotides useful for treating e.g. pulmonary vasoconstriction, inflammation, allergies, asthma, hypertension, bronchitis, emphysema, respiratory distress syndrome, ischemia or cancers

Disclosure: Page 602; 1343pp; English.

The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischemic conditions, pulmonary vasoconstriction, allergies, asthma, impaired respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas, carcinomas, and cancers which may metastasise to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA3992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.

Sequence 2615 BP; 616 A; 763 C; 627 G; 609 T; 0 other;

Alignment Scores:

Pred. No.:	4,32e-137	Length:	2615
Score:	1891.00	Matches:	372
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.73%	Mismatches:	0
Query Match:	99.79%	Indels:	0
DB:	21	Gaps:	0

US-09-836-602-2 (1-373) x AAA34776 (1-2615)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
 DB 294 ATGCCCGGGGCTGGACCGGGTTCCTGCTGAGTTCTGCCCTCTGGGTTCATGAGT 353
 QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 DB 354 CTTGACACACGGTACTGCTACCCAGAGATTACCTACCAGGGACATTTCAATGTT 413

QY	41	SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu	60
DB	414	TTTACAAATGTATCTTACCAAGAAACTACACACCTAGTACCTTGAAGTAGTACCGCTG	473
QY	61	HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys	80
DB	474	CACCCCTGTCTCAACATGGCAATGAGGCCACAAACAAACATCACAGAAACGACAGTCAA	533
QY	81	PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer	100
DB	534	TTTACATCTACCTCTGTGATAACCTCAGTTATGGAAACACAAACTCTTCTGTCAGTCA	593
QY	101	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
DB	594	CAGACCTCTGTAATCAGCACAGTGTTCACACCCAGCCAAACGTTTCAACTCCAGAGACA	653
QY	121	ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer	140
DB	654	ACCTTTGAAGCCTAGCCTGTCACTGGAAATGTTTCAGACCTTTCAACCACTAGCACTAGC	713
QY	141	LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys	160
DB	714	CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCCTATCCTAAGTGACATCAAG	773
QY	161	AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu	180
DB	774	GCAGAAATCAATGTTTCAGGCATCAGAGAAGTGAAATTCACTCAGGGCATCTGCCTGGAG	833
QY	181	GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg	200
DB	834	CAAAATAAGACCTCCACAGCTGTGCGGAGTTTAAAGAAGCAGAGGGAGGCGCTGGCCCA	893
QY	201	ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyValGlnValCysSerLeuLeu	220
DB	894	GTGCTGTGTGGGAGGAGCAGGCTGATGCTGATGTGGGGCCCGCAGGTATGCTCCCTGCTC	953
QY	221	LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu	240
DB	954	CTTGCCACGCTCTGAGGTGAGGCTCAGTCTACTCTGCTGTGGTCTTGGCCCAACAGACAGAA	1013
QY	241	IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle	260
DB	1014	ATTTCCAGCAAACTCCAACTTATGAAAAAGCACCAATCTGACCTGAAAAAGCTGGGGATC	1073
QY	261	LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle	280
DB	1074	CTAGATTTCACTGAGCAAGATGTTGCAACCCAGCAGAGCTATTTCCAAAAGACCTTGATT	1133
QY	281	AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet	300
DB	1134	GCACGTGTCACCTCGGGAGCCCTGCTGGCTGCTTGGGCATCACTGGCTATTTCTCTGATG	1193
QY	301	AsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyrThrThrGlu	320
DB	1194	AATCGCCGCGAGCTGGAGCCCCACAGGAGAAAGCTGGGCGAAGACCTTATTACACGGAA	1253
QY	321	AsnGlyGlyGlnGlnTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys	340
DB	1254	AACGGTGGAGGCCAGGGCTATAGCTCAGGACCTGGGACCTCCCTCCTGAGGCTCAGGAAAG	1313
QY	341	AlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGly	360
DB	1314	GCCAGTGTGAACCGAGGGGCTCAGAAAAACGGGACCGGCGGACCCACCTCCAGAAACGGC	1373
QY	361	HisSerAlaArgGlnHisValValAlaAspThrGluLeu	373
DB	1374	CATTACGAAGACACACACGTGGTGTGCTGATACGAATTG	1412
RESULT	5		
ID	AAF20899		
XX	AAF20899	standard; DNA; 3490 BP.	
AC	AAF20899;		

XX 14-MAR-2001 (first entry)
 XX Human CD-34 polynucleotide fragment #2466.
 DE
 XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KW immunosuppressive; antilasthmatic; analgesic; hypotensive; cytostatic;
 KW respiratory obstruction; pulmonary obstruction; impeded respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200062736-A2.
 PN
 XX
 XX 26-OCT-2000.
 PD
 XX
 XX 24-MAR-2000; 2000WO-US08020.
 PF
 XX
 XX 06-APR-1999; 99US-0127958.
 PR
 XX (UVEC-) UNIV EAST CAROLINA.
 PA (NYCE/) NYCE J W.
 PA
 XX Nyce JW;
 PI
 XX WPI; 2000-679539/66.
 DR
 XX
 XX Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -
 XX
 XX Disclosure; Page 291-292; 1592pp; English.
 PS
 XX The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antilasthmatic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impeded respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.
 XX
 SQ Sequence 3490 BP; 880 A; 966 C; 872 G; 772 T; 0 other;
 Alignment Scores:
 Pred. No.: 6.34e-137 Length: 3490

Score: 1891.00 Matches: 372
 Percent Similarity: 100.00% Conservativity: 1
 Best Local Similarity: 99.73% Mismatches: 0
 Query Match: 99.79% Indels: 0
 DB: 21 Gaps: 0
 US-09-836-602-2 (1-373) x AAF20899 (1-3490)
 Qy 1 MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
 Db 1169 ATGCCCGGGGCTGGACCGCGCTTGTGCTGAGTTTGCTTGGCTTCGGGTTCATGAGT 1228
 Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 Db 1229 CTTGACAAACACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAATGTT 1288
 Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
 Db 1289 TCTACAAATGTATCTTACCAAGAACTACAACACCTAGTACCTTGGAAAGTACCAGCCTG 1348
 Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
 Db 1349 CACCTGTGTCTCAACATGGCAATGAGGCCACAAACATACACAGAAACACAGTCAAA 1408
 Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 Db 1409 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTTGTGTCCAGTCA 1468
 Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
 Db 1469 CAGACCTCTGTAATCAGCACAGTGTTCACACCCCGCAGCAACGTTTCAACTCCAGAGACA 1528
 Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
 Db 1529 ACCTTGAAGCCCTAGCTGTCCACCTGGAAATGTTTCAGACCTTTCACCACTAGCACTAGC 1588
 Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
 Db 1589 CTTGCAACATCTCCACTAAACCCCTATACATCATCTTCTCTATCTAAGTGACATCAAG 1648
 Qy 161 AlaGluIleLysCysSerGlyIleArgGluVallysLeuThrGlnGlyIleCysLeuGlu 180
 Db 1649 GCAGAAATCAATGTTCAGGCATCAGAGAGTGAATTTGACTCAGGGCATCTGCTGGAG 1708
 Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
 Db 1709 CAATTAAGACCTCCAGCTGTCCGAGTTTAAAGAGACAGGGAGAGGGCTGGCCCGA 1768
 Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
 Db 1769 GTGCTGTGTGGGAGGAGCAGCTGATCTGCTGCGGGCCAGGTATGCTCCCTGCTC 1828
 Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThrGlu 240
 Db 1829 CTTGCCAGTCTGAGGTGAGGCCCTCAGTGTCTACTGCTGTGCTTGGCCCAACAGACAGAA 1888
 Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLysGlyIle 260
 Db 1889 ATTTCCAGCAACCTCCAACTTATGAAAAGCACCACCAATCTGACCTGAAAAGACTGGGGATC 1948
 Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
 Db 1949 CTAGATTTCAGTACGACAGATGTTTGAAGCCACAGAGCTATTCCCAAAAGACCTGATT 2008
 Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
 Db 2009 GCACTGTGTCACCTCGGGGCCCTGCTGGCTGTCTTGGGCATCACTGGCTATTTCCTGATG 2068
 Qy 301 AsnArgArgSerTyrProThrGlyCysGluArgLeuGlyGluAspProTyrTyrThrGlu 320
 Db 2069 AATGCCGACCTGGAGCCCCACAGGAAAGGCTGGCGAAGACCCCTATTATACAGGAA 2128
 Qy 321 AsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys 340

|||||
Db 2129 AACGGTGAGGCGGCGGTATAGCTCAGGACCTCGGACCTCCCTCGAGGCTCAGGGAAG 2188
Qy 341 AlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGly 360
Db 2189 GCCAGTGTGAAACCGAGGCGCTCAGAAAAACGGGACCGCCGCGGACCTCCAGAAACGGC 2248
Qy 361 HisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
Db 2249 CATTGAGCAGACACACGCTGGTGGCTGATACCGAATGG 2287
RESULT 6
AAA34777
ID AAA34777 standard; DNA; 3490 BP.
XX
XX
AC AAA34777;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2466.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
XX WO200009525-A2.
PN
XX
PD 24-FEB-2000.
XX
XX 03-AUG-1999; 99WO-US17712.
PF
XX
PR 03-AUG-1998; 98US-0095212.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure; Page 603; 1343pp; English.
XX
CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasize to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences

CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
XX given in the sequence listing.
SQ Sequence 3490 BP; 880 A; 966 C; 872 G; 772 T; 0 other;
Alignment Scores:
Pred. No.: 6.34e-137 Length: 3490
Score: 1891.00 Matches: 372
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.73% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 21 Gaps: 0
US-09-836-602-2 (1-373) x AAA34777 (1-3490)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 1169 ATGCCGCGGGCTGGACCGCGCTTTGCTGCTGAGTTGCTGCCCTCTGGGTTCAGT 1228
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 1229 CTTTGACAAACAGCGTACTGCTACCCGAGAGTTACCTACCCAGGGAACATTTTCAATGTT 1288
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 1289 TCTCAAAATGATCTCTACCAAGAACTACACACCTAGTACCTTGGAGTAGTACCGCTCG 1348
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVallys 80
Db 1349 CACCCTGTGCTCAACATGGCAATGAGGCCACACAAACATCACAGAACGACAGTCAAA 1408
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 1409 TTCACATCTACCTCTGTGATACCTCAGTTTATGGAACACAAACTCTTCTGTCCAGTCA 1468
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 1469 CAGACCTCTGTAAATCAGCACAGTGTTCACCACCCAGCCCAACGTTTCACTCCAGAGACA 1528
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 1529 ACCTTGAAGCCTAGCCTGTCACTTGAAATGTTTTCAGACCTTTCAACCACTAGCAGTACG 1588
Qy 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 1589 CTTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTTAAGTACATCAAG 1648
Qy 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 1649 GCGGAAATCAATGTTTCAGGCATCAGAGAAAGTGAATGACTCAGGGCATCTGCCCTGGAG 1708
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlyLeuAlaArg 200
Db 1709 CAAAATAAGACCTCCAGCTGTGCGGAGTTTAAAGAGGACAGGGAGAGGCGCTGGCCCA 1768
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 1769 GTGCTGTGTGGGAGGAGCAGCTGATGCTGATGTGGGGCCCAAGTATGCTCCCTGTCTC 1828
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 1829 CTTGCCAGTCTGAGTGGAGGCTCAGTGTCTACTGTCTTGGCCCAACAGACAGAA 1888
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLeuGlyIle 260
Db 1889 ATTTCCAGCAAACTCCAACTTATGAAAGACCAATCTGACCTGAAAGAGCTGGGGATC 1948
Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 1949 CTAGATTTTCACTGAGCAAGATGTGTGCAAGCCACAGAGCTATTCCCAAGAGACCTGATT 2008

QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
 |||||
 Db 2009 GCACCTGGTCACTCGGAGCCCTGGCTGTCTTGGGCATCACTGGCTATTTCCCTGATG 2068
 |||||
 QY 301 AsnArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyrThrGlu 320
 |||||
 Db 2069 AATCGCGCAGCTGGAGCCGCCACAGAGAAAGGCTGGGGAGACCCCTTATTACAGGAA 2128
 |||||
 QY 321 AsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys 340
 |||||
 Db 2129 AACGGTGGAGGCCAGGGCTATAGCTCAGGACCTGGGACCTGCCCTCAGGCTCAGGGAAG 2188
 |||||
 QY 341 AlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGly 360
 |||||
 Db 2189 GCCAGTGTGAACCGAGGGCTCAGAAAAGGGAGCGCCAGCGCCACCTCCACAAACGGC 2248
 |||||
 QY 361 HisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
 |||||
 Db 2249 CATTCAGCAAGACACACGCTGGTGGCTGATACCGAATTG 2287
 |||||

RESULT 7
 AAD22662
 ID AAD22662 standard; DNA; 951 BP.
 XX AC AAD22662;
 DT 26-FEB-2002 (first entry)
 XX DE Human CD34 surface antigen truncated variant (tCD34) encoding DNA.
 XX HU Human; surface marker; surface antigen; T lymphocyte; gene therapy;
 KW CD34 truncated variant; tCD34; ds.
 XX OS Homo sapiens.
 OS Synthetic.
 XX Key
 FH 1.951
 FT CDS
 FT /tag- a
 FT /product- "Human CD34 surface antigen
 FT truncated variant (tCD34)"
 XX EP1148066-A1.
 XX PD 24-OCT-2001.
 XX PF 18-APR-2001; 2001EP-0109374.
 XX PR 18-APR-2000; 2000DE-1019075.
 XX PA (ZAND/) ZANDER A R.
 XX Zander AR;
 XX WPI; 2002-019289/03.
 XX P-PSDB; AAEL3542.
 XX New gene transfer vector (accession number DSM13396) containing a
 PT transgene and a nucleic acid sequence coding for a surface marker,
 PT useful in gene therapy, and for detecting genetically modified cells or
 PT cells which do not express CD34 -
 XX
 PS Claim 3; Page 12-13; 28pp; English.
 XX The patent discloses a gene transfer vector (accession number DSM13396)
 CC containing a transgene and a nucleic acid sequence coding for a surface
 CC marker. The surface marker is the CD34 surface antigen, its fragment or
 CC variant. The vector is useful for in vitro transduction of T lymphocytes,
 CC for gene therapy, and in the enrichment, detection and analysis of cells
 CC in vitro that do not naturally express CD34. T lymphocytes transduced
 CC with the vector are also useful in gene therapy. The CD34 nucleic acid
 CC sequences (marker genes), their fragments or variants are used for

CC detecting genetically modified cells or cells which do not naturally
 CC express CD34. The present sequence is a DNA encoding human CD34 surface
 CC antigen truncated variant (tCD34).
 XX
 SQ Sequence 951 BP; 261 A; 264 C; 213 G; 213 T; 0 other;

Alignment Scores:
 Pred. No.: 4,26e-113 Length: 951
 Score: 1574.00 Matches: 313
 Percent Similarity: 99.37% Conservative: 1
 Best Local Similarity: 99.05% Mismatches: 2
 Query Match: 83.06% Indels: 0
 DB: 24 Gaps: 0

US-09-836-602-2 (1-373) x AAD22662 (1-951)

QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
 |||||
 Db 1 ATGCCGCGGGCTGGACCGGCTTGTCTGTGAGTTGCTTCTGGGTTTCATGAGT 60
 |||||
 QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
 |||||
 Db 61 CTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 120
 |||||
 QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
 |||||
 Db 121 TCTACAAATGTATCTTACCAAGAACTACAAACACCTAGTACCTTGGAGTACGAGCTG 180
 |||||
 QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
 |||||
 Db 181 CACCCCTGTGTCTCAACATGGCAATGAGGCCACCAACAACATCACAGAAACGACAGTCAA 240
 |||||
 QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
 |||||
 Db 241 TTCACATCTACCTCTCTGTGATACTCAGTTTATGGAACAACAACACTCTCTGTCCAGTCA 300
 |||||
 QY 101 GlnThrSerValIleSerThrValPheThrProAlaAsnValSerThrProGluThr 120
 |||||
 Db 301 CAGACCTCTGTAATCAGCACAGTGTTCACACCCCAACGTTTCACTCCAGAGACA 360
 |||||
 QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
 |||||
 Db 361 ACCTTGAAGCCTAGCCTGTCCCTGGAAATGTTTCAGACCTTTCACCCACTAGCCTAGC 420
 |||||
 QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLys 160
 |||||
 Db 421 CTTGCAACATCTCCACTAACCCATACATCATCTTCTCTATCTTAAGTACATCAAG 480
 |||||
 QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
 |||||
 Db 481 GCAGAAATCAATGTTTCAGGCATCAGAGAAAGTGAATTTGACTCAGGGCATCTGCCTGGAG 540
 |||||
 QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
 |||||
 Db 541 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAGACAGGGGAGAGGCGCTGGCCCGA 600
 |||||
 QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAlaGlyAlaGlnValCysSerLeuLeu 220
 |||||
 Db 601 GTCTGTGTGGGAGGAGCAGGCTGATGCTGGGCCCCAGGATGCTCCCTCCTC 660
 |||||
 QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
 |||||
 Db 661 CTTGCCCAAGTCTGAGGTGAGGCCCTCAGTGTCTACTGTGCTTGTGGCAACAGAACAGAA 720
 |||||
 QY 241 IleSerSerLysLeuGlnLeuMetLysHisGlnSerAspLeuLysLeuGlyLe 260
 |||||
 Db 721 ATTTCAGCAAACTCCAACTTATGAAAAGACCAACTCTGACCTGAAAAGAGCTGGGATC 780
 |||||
 QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
 |||||
 Db 781 CTAGATTTCAGTACGCAAGATGTTGCAAGCCACAGAGCTATTCCCAAAAAGACCTGATT 840
 |||||
 QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
 |||||

Db 841 GCACCTGTCACCTCGGGAGCCCTGCTGGCTGTCTTGGCATCACCTGGCTATTTCCTGATG 900
QY 301 AsnArgSerTrpSerProThrGlyGluArgLeuGlyGluAspPro 316
Db 901 AATCGCCGAGCTGGAGCCCCACAGGAGAAAGCTGGAACCTAGAACCA 948
RESULT 8
AAD22663
ID AAD22663 standard; DNA; 906 BP.
XX AAD22663;
AC AAD22663;
XX
DT 26-FEB-2002 (first entry)
XX Human CD34 surface antigen deleted variant (dCD34) encoding DNA.
DE Human; surface marker; surface antigen; T lymphocyte; gene therapy;
KW CD34 deleted variant; dCD34; ds.
XX Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..906
FT /tag= a
FT /product= "Human CD34 surface antigen
FT deleted variant (dCD34)"
XX
PN EP1148066-A1.
XX
PD 24-OCT-2001.
XX
PF 18-APR-2001; 2001EP-0109374.
XX
PR 18-APR-2000; 2000DE-1019075.
XX
PA (ZAND/) ZANDER A R.
XX
PI Zander AR;
XX
DR WPI: 2002-019289/03.
DR P-PSDB; AAEE13543.
XX
XX New gene transfer vector (accession number DSM13396) containing a
PT transgene and a nucleic acid sequence coding for a surface marker,
PT useful in gene therapy, and for detecting genetically modified cells or
PT cells which do not express CD34 -
XX
PS Claim 3; Page 14-15; 28pp; English.
XX
CC The patent discloses a gene transfer vector (accession number DSM13396)
CC containing a transgene and a nucleic acid sequence coding for a surface
CC marker. The surface marker is the CD34 surface antigen, its fragment or
CC variant. The vector is useful for in vitro transduction of T lymphocytes,
CC for gene therapy, and in the enrichment, detection and analysis of cells
CC in vitro that do not naturally express CD34. T lymphocytes transduced
CC with the vector are also useful in gene therapy. The CD34 nucleic acid
CC sequences (marker genes), their fragments or variants are used for
CC detecting genetically modified cells or cells which do not naturally
CC express CD34. The present sequence is a DNA encoding human CD34 surface
CC antigen deleted variant (dCD34).
XX
SQ Sequence 906 BP; 247 A; 250 C; 199 G; 210 T; 0 other;

Alignment Scores:
Pred. No.: 4.37e-108 Length: 906
Score: 1509.00 Matches: 301
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 79.63% Indels: 0
DB: 24 Gaps: 0

US-09-836-602-2 (1-373) x AAD22663 (1-906)
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 1 ATGCCCGGGGCTGGACCGCTTCTTGGCTGAGTTTCTGCTTCTGGTTTCATGAGT 60
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 61 CTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAATGTT 120
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 121 TCTACAAATGTATCTTACCAAGAACTACAAACACCTAGTACCTTGGAGTACCAAGCTG 180
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 181 CACCCTGTCTCAGATGGCAATGAGGCCACAAACAACATCACAGAAACGACAGTCAAA 240
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 241 TTCACATCTACCTCTCTGTATACCTCAGTTTATGGAACACACAAACTCTTCTCTCCAGTCA 300
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 301 CAGACCTCTGTAATCAGCACAGTGTTCACCAACCCAGCCCAACGTTTCACTCCAGAGACA 360
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 361 ACCTTGAAGCCTAGCCTGTCACTGGAATGTTTCAGACCTTTCACCACTACACTAGCTAGC 420
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 421 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTTAAGTGACATCAAG 480
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 481 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTTGACTCAGGGGCATCTCGCTGGAG 540
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 541 CAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAGGACAGGGGAGGGGCTTGGCCCGA 600
QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 601 GTGCTGTGGGGAGGAGCAGGCTGATGCTGTGCTGGGCCCCAGGTATGCTCCCTGCTC 660
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 661 CTTGCCACGTCTCAGGTGAGGCTCAGTGTCTACTGTGTCTTGGCCAAACAGAACAGAA 720
QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLeuGlyIle 260
Db 721 ATTTCAGCAAACTCCAACTTATGAAAGACCACTATGACCTGAAAAAGCTGGGGATC 780
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 781 CTAGATTCTCACTCAGCAAGATGTTGCAAGCCACAGAGCTATTCCTCCAAAGACCTGATT 840
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 841 GCACCTGTCACTCGGAGGCCCTGCTGGCTGTCTTGGGCATCATCTGGCTATTTCCTCATG 900
QY 301 Asn 301
Db 901 AAT 903
RESULT 9
AAT59508
ID AAT59508 standard; cDNA; 1065 BP.
XX
AC AAT59508;
XX
DT 27-OCT-1997 (first entry)

XX Porcine CD34 from clone LAX2-3 encoding cDNA.
 XX
 KW Antibody; bone marrow chimera; graft; antigen; pig; immune response;
 KW haematopoietic progenitor cell; stem cell; cord blood; ss.
 XX
 OS Sus scrofa.
 XX
 FH Key Location/Qualifiers
 FT CDS 31..1065
 FT /*tag= a
 FT /product= CD34
 FT sig_peptide 31..123
 FT /*tag= b
 FT mat_peptide 124..1062
 FT /*tag= c
 XX
 FN W09640244-A1.
 XX
 XX 19-DEC-1996.
 PD
 XX 03-JUN-1996; 96WO-US08340.
 PF
 XX 07-JUN-1995; 95US-0475634.
 PR
 XX (BIOT-) BIOTRANSPLANT INC.
 FA
 XX Hawley RJ, Monroy RL;
 PI
 XX WPI; 1997-108628/10.
 DR
 DR P-PSDB; AAW11822.
 XX
 PT Antibodies selective for CD34 + porcine cells - used to isolate such
 PT cells for generating bone marrow chimera(s) that will accept porcine
 PT grafts, also new porcine CD34 antigen
 XX
 PS Example 1; Fig 1; 35pp; English.
 XX
 CC The present sequence encodes porcine CD34. Antibodies (Ab) have been
 CC produced that recognise porcine CD34+ cells. The Ab's are used to
 CC recover CD34+ cells, preferably from porcine bone marrow but also from
 CC cord blood, and CD34+ cells, which are enriched in haematopoietic
 CC progenitor cells, especially stem cells, are used to generate bone
 CC marrow chimeras in a human to assist acceptance of a porcine graft
 CC (i.e. to prevent or inhibit an immune response against the graft).
 CC Ab's can also be used to detect CD34+ cells, e.g. to determine if
 CC mixed chimerism has been established. The CD34 polypeptide and its
 CC fragments are used to raise Ab.
 XX
 SQ Sequence 1065 BP; 248 A; 325 C; 273 G; 219 T; 0 other;
 Alignment Scores:
 Pred. No.: 4.34e-61 Length: 1065
 Score: 904.00 Matches: 196
 Percent Similarity: 69.44% Conservative: 38
 Best Local Similarity: 58.16% Mismatches: 77
 Query Match: 47.70% Indels: 26
 DB: 18 Gaps: 6

US-09-836-602-2 (1-373) x AAT59508 (1-1065)

Qy 1 MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
 |||||
 Db 67 ATGCCGGGGGTTGGACCACTCTGCTGCTGAGTTGCTGCCCTCTGGGTTTCACAGCT 126
 |||||
 Qy 21 LeuAspAsnAnGlyThrAlaThrProGluLeuPro----- 32
 ::|||
 Db 127 GTG---AACAGCTCAACTATTGCTTCCACCTTGGCAGCTGCCGTGGGTCAACTCCCAACC 183
 |||||
 Qy 33 -----ThrGlnGlyThrPheSerAsn----- 39
 |||||
 Db 184 GGGCCGGCTACCGGGGCGAGCTATCACCAGGTCACATATCTCAGACATATCTTCACCT 243

Qy 40 ValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSer 59
 |||||
 Db 244 GTTCTCAATATATATCAACGAGAAACACATCA---GATGCTTTTGAAGTCCACAGC 300
 |||||
 Qy 60 LeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVal 79
 |||||
 Db 301 CTCCACACTGCTCTCAGGCGAGCAGTGGGACCACTGAGCATCTCAGGCCCTACAGTT 360
 |||||
 Qy 80 LysPheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGln 99
 |||||
 Db 361 AATTTCATGCTACTCTGGCGGTCACCTCGTCCCGAAACCGTTAACTCTCTGTCTCCAG 420
 |||||
 Qy 100 SerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGlu 119
 |||||
 Db 421 CCTCAGACCTCT---CTAGCCACAGCGTCTCCGCCACCATCAACTTTACAACCTCAGAG 477
 |||||
 Qy 120 ThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThr 139
 |||||
 Db 478 GTGACCTCTGCAGCCACGACGCTTCCAGGAATGTTTCCAGACCCCTCTACAACAGTACC 537
 |||||
 Qy 140 SerLeuAlaThrSerProThrLysProThrThrSerSerSerProIleLeuSerAspIle 159
 |||||
 Db 538 AGCCTGCGAGATCCCCACCGCCCTACACATCATCTCTCTTACCCTCCAGGTAGCCAC 597
 |||||
 Qy 160 LysAlaGluIleLysCysSerGlyLeuArgGluValLysLeuThrGlnGlyLeuCysLeu 179
 |||||
 Db 598 AAGGGGAAGTCAATGTGCCCAATCAACAGAGTGAATGACCCAGGTATCTGCGCTG 657
 |||||
 Qy 180 GluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla 199
 |||||
 Db 658 CAGCGAATGAGACTCGCGTGGGAGAGTAAAGAGGACAATGGAGAGAGTGTGATG 717
 |||||
 Qy 200 ArgValLeuCysGlyGluGluGlnAlaAspAlaGlyAlaGlnValCysSerLeu 219
 |||||
 Db 718 CAATCTGTGTGGCAGGAGGAGCTGAGCGCGGCGCAGGG-----GTGTGCTCTCTG 771
 |||||
 Qy 220 LeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuLeuValLeuAlaAsnArgThr 239
 |||||
 Db 772 CTCCTTGCCCAATCTGAGGTGAACCTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 831
 |||||
 Qy 240 GluLeuSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGly 259
 |||||
 Db 832 GAACCTTAGCACAAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 891
 |||||
 Qy 260 IleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeu 279
 |||||
 Db 892 ATCCAAACTTCTCGAAACAAGATGTTAGGAGCCACAGAGCTACTCCCGAAACCTTG 951
 |||||
 Qy 280 IleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeu 299
 |||||
 Db 952 ATTGCACTGGTCACTCGGGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1011
 |||||
 Qy 300 MetAsnArgArgSerTyrProThrGlyCysLeuGlyGluAspPro 316
 |||||
 Db 1012 ATGAACCGTCGACTGGAGCCCTACAGGAAAGGCTGGAGCTGGAACCC 1062
 |||||
 RESULT 10
 AAC00249
 ID AAC00249 standard; cDNA; 615 BP.
 XX AAC00249;
 AC AAC00249;
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 247.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX

PD 06-SEP-2000.
XX
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
XX PR 26-FEB-1999; 99US-0122487.
XX
XX PA (GEST) GENSET.
XX
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
DR P-PSDB; AAG00243.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX
PS Claim 1; SEQ ID 247; 7lpp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 615 BP; 149 A; 181 C; 159 G; 119 T; 7 other;

Alignment Scores:
Pred. No.: 1-05e-30 Length: 615
Score: 508.00 Matches: 99
Percent Similarity: 96.12% Conservative: 0
Best Local Similarity: 96.12% Mismatches: 4
Query Match: 26.81% Indels: 0
DB: 21 Gaps: 0
US-09-836-602-2 (1-373) x AAC00249 (1-615)
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
DB 305 ATGCCGCGGGCTGGACCGCGCTTTGCTGCTGAGTTTGCCTTCTGGGTTTCATGAGT 364
QY 21 LeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
DB 365 CTTGACAAACAGGTACTGCTACCCAGAGTTACCTTACCAGGGGACATTTTCAAATGTT 424
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
DB 425 TCTACAAATGTATCTTACCAAGAACTACACACCTTGTCCTTGGAGTACACCCCTG 484
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
DB 485 CACCCCTGTCTCAACATGGCAATGAGGCCACAAACAAATCATCACAGAAACGACAGTSSNG 544
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
DB 545 TTCACATCTASCTCTGTGATAASCTCAGTTTATGGRAACACAAACACTCTTCTGTCCAGTCA 604
QY 101 GlnThrSer 103
DB 605 CAGACCTCT 613
RESULT 11
ABA51134/c
ID ABA51134 standard; DNA; 233 BP.

XX ABA51134;
XX
XX DT 01-FEB-2002 (first entry)
XX
XX DE Human breast cell single exon nucleic acid probe #9829.
XX
XX KW Human; microarray; single exon probe; gene expression; breast;
XX KW disease; cancer; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157271-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00662.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
XX PT New spatially-addressable set of single exon nucleic acid probes,
XX PT useful for measuring gene expression in sample derived from human
XX PT breast, comprises number of single exon nucleic acid probes -
XX
XX PS Claim 4; SEQ ID NO 9829; 327pp + sequence listing; English.
XX
XX CC The invention relates to a spatially-addressable set of single exon
XX CC nucleic acid probes for measuring gene expression in a sample derived
XX CC from human breast and BT 474 cells. The method involves contacting
XX CC the probes with a collection of detectably labelled nucleic acids
XX CC derived from mRNA of human breast, and then measuring the label
XX CC bound to each probe of the microarray. The probes are useful for
XX CC verifying the expression of regions of genomic DNA predicted to
XX CC encode proteins. They are useful for gene discovery, and for
XX CC determining predisposition and/or prognosing breast disease. Gene
XX CC expression analysis is useful for assessing the toxicity of chemical
XX CC agents on cells. The microarray of this invention presents a far greater
XX CC diversity of probes for measuring gene expression, with far less bias
XX CC than expressed sequence tag microarrays. The method is suitable for
XX CC rapid production of functional information from genomic sequence. The
XX CC present sequence is a single exon nucleic acid probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
Pred. No.: 5.45e-16 Length: 233
Score: 311.00 Matches: 60
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.41% Indels: 0
DB: 22 Gaps: 0
US-09-836-602-2 (1-373) x ABA51134 (1-233)

QY 16 SerGlyPheMetSerLeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGly 35
DB 231 TCTGGGTTTCATGAGTCTTGACAAACAGCGTACTGCTACCCAGAGTTACTACCCAGGGA 172

Qy 36 ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeu 55
|||||
Db 171 ACATTTTCAAAATGTTCTACAAATGTATCCTACCAAGAACTACAAACCTAGTACCCTT 112
Qy 56 GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr 75
|||||
Db 111 GGAAGTACCAGCCTGCACCCTGTGTCTCAACATGGCAATGAGGCCAACAAACATCACA 52

RESULT 12
ABA69132/C
ID ABA69132 standard; DNA: 233 BP.
XX AC ABA69132;
XX DT 01-FEB-2002 (first entry)
XX DE Human foetal liver single exon nucleic acid probe #17437.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX PS Claim 4; SEQ ID NO 17437; 639pp + sequence listing; English.
XX CC The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
Pred. NO.: 5.45e-16 Length: 233
Score: 311.00 Matches: 60
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.41% Indels: 0
DB: 22 Gaps: 0

US-09-836-602-2 (1-373) x ABA69132 (1-233)

Qy 16 SerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeuProThrGlnGly 35
|||||
Db 231 TCTGGGTTCATGAGTCTTGACAAACAGGTACTGCTACCCAGAGTTACCTACCCAGGGA 172

Qy 36 ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeu 55
|||||
Db 171 ACATTTTCAAAATGTTCTACAAATGTATCCTACCAAGAACTACAAACCTAGTACCCTT 112
Qy 56 GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr 75
|||||
Db 111 GGAAGTACCAGCCTGCACCCTGTGTCTCAACATGGCAATGAGGCCAACAAACATCACA 52

RESULT 13
ABA36064/C
ID ABA36064 standard; DNA: 233 BP.
XX AC ABA36064;
XX DT 23-JAN-2002 (first entry)
XX DE Probe #14530 for gene expression analysis in human heart cell sample.
XX KW Human; gene expression; heart; microarray; vascular system; probe;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease; ss.
XX OS Homo sapiens.
XX PN WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00666.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX PS Claim 4; SEQ ID No 14530; 530pp; English.
XX CC The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX CC Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 233 BP; 55 A; 47 C; 61 G; 70 T; 0 other;

Alignment Scores:
Pred. NO.: 5.45e-16 Length: 233
Score: 311.00 Matches: 60
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.41% Indels: 0
DB: 22 Gaps: 0

US-09-836-602-2 (1-373) x ABA36064 (1-233)

Db 231 TCTGGGTTTCATGAGTCTTGACAAACGGTACTGTACCCAGAGTTACCTACCCAGGGA 172
Qy 36 ThrPheSerAsnValSerThrAsnValSerTyrGlnGluThrThrProSerThrLeu 55
Db 171 ACATTTCAAATGTTTCTACAAATGTATCTACCAAGAACTACAACACCTAGTACCCTT 112
Qy 56 GlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThr 75
Db 111 GGAAGTACCAGCCTGCACCCCTGTGTCTCAACATGGCAATGAGGCCACACAAACATCACA 52

Search completed: October 30, 2002, 08:25:37
Job time : 178.922 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 30, 2002, 07:08:08 ; Search time 1912.1 Seconds
(without alignments)
4082.221 Million cell updates/sec

Title: US-09-836-602-2
Perfect score: 1895
Sequence: 1 MPRGWTALCLLSLLPSGFM.....QATSRNGHSARQHVADTEL 373

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n_model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09836602/runat_29102002_101144_3785/app_query.fasta_1.1429
-DB=GenEmbl -FAST=Fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09836602@cgn.1.1.3800 @runat_29102002_101144_3785 -NCPU=6 -TCPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1895	100.0	2463	9	S53910	S53910 CD34-glycop
2	1891	99.8	2615	9	HUMCD34HS	M81104 Human CD34
3	1827	96.4	2657	6	AX333753	AX333753 Sequence
4	1827	96.4	2657	9	S53911	S53911 CD34-glycop
5	1272	67.1	2956	4	CFU49457	U49457 Canis famli
6	1212	64.0	2690	4	AF461503	AF461503 Sus scrof
7	1173.5	61.9	1149	10	S69299	S69299 CD34-cell s
8	1173.5	61.9	1260	10	S69293	S69293 CD34-stem c
9	1173.5	61.9	2398	10	BC006607	BC006607 Mus muscu
10	1061.5	56.0	2453	4	AB021662	AB021662 Bos tauru
11	1019.5	53.8	2420	4	AF259378	AF259378 Capra hir
12	980	51.7	581	9	AF202879	AF202879 Homo sapi
13	904	47.7	1065	6	AR078454	AR078454 Sequence
14	870.5	45.9	978	10	S69301	S69301 CD34-cell s
15	630.5	33.3	160771	9	HS88L2	AL035091 Human DNA
16	630.5	33.3	182742	2	AC084393	AC084393 Homo sapi
17	617	32.6	185162	2	AL365178	AL365178 Homo sapi
18	617	32.6	212535	2	AL356275	AL356275 Homo sapi
19	554	29.2	98347	2	AL513203	AL513203 Mus muscu
20	554	29.2	168333	2	AL513470	AL513470 Mus muscu
21	409	21.6	254	9	HUMCD34S3	M81940 Human CD34
22	325	17.2	1580	11	G06650	G06650 human STS W
23	321	16.9	1386	9	HUMCD34S8	M81945 Human CD34
24	311	16.4	183	9	HUMCD34S2	M81939 Human CD34
25	307	16.2	1330	10	S69295S2	S69302 stem cell a
26	275	14.5	165	9	HUMCD34S7	M81944 Human CD34
27	260	13.7	157	9	HUMCD34S5	M81942 Human CD34
28	233	12.3	235	6	AX247364	AX247364 Sequence
29	188	9.9	3124	5	MMTHRAD1	Y13978 Gallus gall
30	168.5	8.9	12412	3	CELH43E16	AL591180 Zebrafish
31	167	8.8	80272	5	AL591180	AF109393 Rattus no
32	166.5	8.8	5309	10	AF109393	AL591175 Zebrafish
33	166	8.8	89232	5	AL591175	U11583 Saccharomyc
34	162.5	8.6	55069	8	YSGH9196	U20824 Equine herp
35	162.5	8.6	184427	14	EHVU20824	AL669823 Mus muscu
36	160.5	8.5	253273	2	AL669823	AB020726 Rattus no
37	159.5	8.4	1731	10	AB020726	AC019041 Homo sapi
38	158.5	8.4	183358	9	AC019041	L02115 Frog Integu
39	156.5	8.3	2161	5	XELFMC1X	AC007263 Homo sapi
40	155.5	8.2	167390	9	AC007263	AC074046 Mus muscu
41	155.5	8.2	182534	10	AC074046	Z71659 S.cerevisia
42	154	8.1	3053	8	SCVNR044W	M60590 S.cerevisia
43	154	8.1	3137	8	YSCNAGLCS	AL357093 Human chr
44	153.5	8.1	167254	9	CNS050TDS	AC104020 Homo sapi
45	153.5	8.1	172789	2	AC104020	

ALIGNMENTS

RESULT 1
S53910
LOCUS S53910 2463 bp mRNA linear PRI 08-MAY-1993
DEFINITION CD34-glycoprotein expressed in lymphohematopoietic progenitor cells
(alternatively spliced) [human, UT7 cells, mRNA, 2463 nt].
ACCESSION S53910
VERSION S53910.1 GI:264766
KEYWORDS human UT7 cells.

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Nakamura,Y., Komano,H. and Nakauchi,H.
TITLE Two alternative forms of cDNA encoding CD34
JOURNAL Exp. Hematol. 21 (2), 236-242 (1993)
MEDLINE 93146100
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbs 124007] from the original journal article.
This sequence comes from Fig. 1A.
FEATURES
source
1..2463
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene
1..2463
CDS
91..1248
/gene="CD34"
/translation="MLVRRCARAGPRMGWATLCLLSLPSGFMSLDNNGTATPELP
TQTFESNVSNVSYQETTPSTLGSSTSLHPVSOHNEATTNITETVTSTVITSY
YGNVNSVQSOTSIVSTVFTTPANVSTPETTLKPSLSPGNVSDLTSTSLATSPTRP
YTSPPSLSDIKAEIKCSGIREVKLTQGLCEONKTSSCAEFKKDRGEGEARLVLCGEE
QADADAGLQSLLEAQSEVRPQCLLVLANRTEISKKLQMKKHSDLKGLIDFT
EQDVASHQSYQSKTLTALVTSGLALLVIGITGYFLMNRNRSWPTGRRLGDPYYTENG
GGQGYSSGPGTSPTEAOGKASVNRGAQNRGQATSRNGHSARQHVVADTEL"
BASE COUNT 589 a 709 c 580 g 585 t
ORIGIN
Alignment Scores:
Pred. No.: 4,45e-115 Length: 2463
Score: 1895.00 Matches: 373
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x 553910 (1-2463)

QY 1 MetProArgGlyTTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 127 ATGCCGGGGGCTGGACCGCCCTTGCTTGCTGAGTTGCTGCTTCTGGGTTTCATGAGT 186

QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 187 CTTGACAACAACGGTACTGCTACCCAGAGTTACCTACCCAGGGAACATTTTCAAAATGTT 245

QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 247 TCTACAATGATCTCTACCAAGAACTACACACCTAGTACCTTGGAAAGTACGAGCGTG 306

QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValIlys 80
Db 307 CACCCTGTGCTCAACATPGGAATGAGCCCAACAACATCACAGAAACACAGTCAAA 366

QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 367 TTCACATCTACCTCTGTGATAACCTCAGTTTATGGAACACAAACTCTCTGTCTCCAGTCA 426

QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 427 CAGACCTCTGTAATCAGCACAGTGTTCACCAACCCAGCCCAACGTTTCAACTCCAGAGACA 486

QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 487 ACCTTGAGCCCTGACCTGCTGGAATGTTTTCAGACCTTTCACCACTAGCAGTACG 546

```

```

QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLys 160
Db 547 CTTGCAACATCTCCACTAAACCTATACATCATCTTCTCTATCTTAAAGTGAATCAAG 606

QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 607 GCAGAAATCAAAATGTTTCAGGCATCAGAGAGTGAATTTGACTCAGGGCATCTGCCTGGAG 666

QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 667 CAAAATAAGACCTCCAGCTGTCGGAGTTTAAAGAGACAGAGGGAGGGGCTGGCCCGA 726

QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 727 GTGCTGTGTTGGGAGGAGCAGGCTGATGCTCTGGGCCCCAGGTATGCTCCTCCTC 786

QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 787 CTTGCCAGCTGTGAGGTGAGGCCCTCAGTGTCTACTGTGCTTGGCCACAGAACAGAA 846

QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 847 ATTTCCAGCAAACTCCAACTTATGAAAAGCACCACCAATCTGACCTGAAAAGCTGGGATC 906

QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 907 CTAGATTTCAGTGAGCAAGATGTTGCAAGCCACAGAGCTATTCCCAAAAGACCTGATT 966

QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 967 GCACCTGCTACCTCGGAGGCCCTGCTGCTCTGTTGGCATCACTGGCTATTTCTCTGATG 1026

QY 301 AsnArgSerTTrpSerProThrGlyGluArgLeuGlyGluAspProTyrTyrThrGlu 320
Db 1027 AATCGCCGAGCTGGAGCCCCACAGAGAAAGGCTGGCGAAGACCTTATTACACGGAA 1086

QY 321 AsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys 340
Db 1087 AACGGTGGAGCCAGGCTATAGCTCAGGACCTGGGACCTCCCTCCTGAGGCTCAGGAAAG 1146

QY 341 AlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGly 360
Db 1147 GCCAGTGTGAACCGAGGGGCTCAGAAACAGGGACCGCCAGCCACCTCCAGAAACGGC 1206

QY 361 HisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
Db 1207 CATTCAACAGACAAACACAGTGGTGGCTGATACCGAATTG 1245

RESULT 2
HUMCD34HS HUMCD34HS 2615 bp mRNA linear PRI 02-NOV-1993
LOCUS Human CD34 mRNA, complete cds.
DEFINITION M81104.60172
ACCESSION M81104.1 GI:180108
VERSION M81104.1 GI:180108
KEYWORDS CD34; hematopoietic stem cell surface antigen; sialomucin.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Simmons,D.L., Satterthwaite,A.B., Tenen,D.G. and Seed,B.
TITLE Molecular cloning of a cDNA encoding CD34, a sialomucin of human
hematopoietic stem cells
JOURNAL J Immunol. 148, 267-271 (1992)
MEDLINE 92091783
FEATURES
source
1..2615
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
CDS
294..1415
/codon_start=1
/product="CD34"
/protein_id="AAA03181.1"

```

/db_xref="GI:180109"
/translation="MRGWTAALCLLSLPSGMSLDNNGTATPELPTQGFSSNVSTNV
SYEETTPSLGSGTSLHPVSHGNEATTNITETVKFTSTSVTSVITGNTNSVQSQT
SVLSTVFTTPANVSTPETTLKPSLSPCNVSLSTTSTSLATSPKPYTSSPILSDIK
AEIKCSGIREVLTGLICLEQNTSSCAEFKDRGRLGVLGEGQADADAGVCS
LLAQSEVRPOCLLLVLANKEISSQLMKKHOSDLKLGILDFEDVDVASHOSYSQ
KTLLALVTSRGLLAVLGITGYFLMNRNSPSTGERLGEDPYTENGGGGYSSTGPGTS
PEAQKASVNRGAQKNGTGQATSRNGHSARQHVVADEL"
BASE COUNT 616 a 763 c 627 g 609 t
ORIGIN

Alignment Scores:

Pred. No.: 8,79e-115 Length: 2615
Score: 1891.00 Matches: 372
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.73% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 9 Gaps: 0

US-09-836-602-2 (1-373) x HUMCD34HS (1-2615)

Qy 1 MetProArgGlyThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 294 ATGCGGGGGCTGGACCGCTTGCTGCTGAGTTGCTGCTTCTGGGTTTCATGAT 353
Qy 21 LeuAspAsnGlyThrAlaLeuProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 354 CTTGACAAACAGCGTACTGCTACCCAGAGTTACTACCCAGGGAACATTTTCAAAATGT 413
Qy 41 SerThrAsnValSerTyrglnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 414 TCTACAAATGATCTCTACCAAGAACTACAAACACTAGTACCTTGGAAAGTACCAGCGTG 473
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80
Db 474 CACCCTGTCTCAACATGCGCAATGAGGCCACAAACATCACAGAAACAGACAGTCAAA 533
Qy 81 PheThrSerThrSerValIleThrSerValTyrglnGlyAsnThrAsnSerSerValGlnSer 100
Db 534 TTCACATCTACTCTGTGATTAACCTCAGTTTATGGAACACAAACTCTTCTGTCAGTCA 593
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 594 CAGACCTCTGTAATCAGCACAGTGTTCACACCCAGCCCAACGTTTCAACTCCAGAGACA 653
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 654 ACCTTGAAGCCTAGCCTGTCACTGGAAATGTTTCAGACCTTTCAACCACTAGCACTAGC 713
Qy 141 LeuAlaThrSerProThrLysProTyrrThrSerSerSerProIleLeuSerAspIleLys 160
Db 714 CTTGCAACATCTCCCACTAAACCTATACATCATCTTCTCTATCTTCTTAAGTGACATCAAG 773
Qy 161 AlaGluLysCysSerGlyIleAargluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 774 GCAGAAATCAAATGTTCAAGGCATCAGAGAAGTGAATGACTCAGGGCATCTGCGTGGAG 833
Qy 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 834 CAAATAGACCTCCAGCTGTCGCGAGTTTAAAGAGGACAGAGGAGAGGGCTGCGCCCGA 893
Qy 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 894 GTGCTGTGGGAGGAGAGCGCTGATGCTGCTACTGCTGGTCTTGGCCAAACAGACAGAA 1013
Qy 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 954 CTTGCCAGTCTGAGGTGAGGCTCAGTGTCTACTGCTGGTCTTGGCCAAACAGACAGAA 1013
Qy 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysGlyIle 260
Db 1014 ATTTCCAGCAACTCCCACTTATGAAAAAGCACCACCTGACCTGAAAAAGCTGGGATC 1073

Qy 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrrSerGlnLysThrLeuIle 280
Db 1074 CTAGATTTCACTGACGAAGATGTTGCAAGCCACAGAGCTATTCCCAAAGACCTGATT 1133
Qy 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrrPheLeuMet 300
Db 1134 GCATGTGTACCTCGGAGCCCTGCTGGCTGCTTGGGCATCACTGGCTATTTCCTGATG 1193
Qy 301 AsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyrrThrGlu 320
Db 1194 ANTCCCGCAGCTGGAGCCCCACAGAGAGAAAGCTGGCGGAGAACCTTATTACACGAA 1253
Qy 321 AsnGlyGlyGlnGlyTyrrSerSerGlyProGlyThrSerProGluAlaGlnGlyLys 340
Db 1254 AACGTTGGAGGCCAGGCTATAGCTCAGGACCTGGACCTCCCTGAGGCTCAGGAAAG 1313
Qy 341 AlaSerValAsnArgGlyAlaGlnGlnAsnGlyThrGlyGlnAlaThrSerArgAsnGly 360
Db 1314 GCCAGTGTGAACCGAGGGGCTCAGAAACCGGACCGCCAGGCCACCTCCAGAAACGGC 1373
Qy 361 HisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
Db 1374 CATTACAGACACACACGCTGCTGGCTGATACCGAATG 1412
RESULT 3
AX333753 2657 bp DNA linear PAT 09-JAN-2002
LOCUS AX333753 Sequence 4262 from Patent WO0194629.
DEFINITION AX333753
ACCESSION AX333753
VERSION AX333753.1 GI:18124472
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 4262 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES Location/Qualifiers
source 1..2657
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 624 a 790 c 609 g 634 t
ORIGIN
Alignment Scores:
Pred. No.: 1.38e-110 Length: 2657
Score: 1827.00 Matches: 372
Percent Similarity: 84.93% Conservative: 0
Best Local Similarity: 84.93% Mismatches: 1
Query Match: 96.41% Indels: 66
DB: 6 Gaps: 1
US-09-836-602-2 (1-373) x AX333753 (1-2657)
Qy 1 MetProArgGlyThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 127 ATGCGGGGGCTGGACCGCTTGCTGCTGAGTTGCTGCTTCTGGGTTTCATGAT 186
Qy 21 LeuAspAsnGlyThrAlaLeuProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 187 CTTGACAAACAGCGTACTGCTTACCCAGAGTTACTACCCAGGGAACATTTTCAAAATGT 246
Qy 41 SerThrAsnValSerTyrglnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 247 TCTCAAAATGATCTCTACCAAGAACTACAAACCTAGTACCTTGGAAAGTACCAGCGCTG 306
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrValLys 80

```
Db 307 CACCCTGTGCTCAACATGGCAATGAGGCCACACAAACATCATCAGAAACGACAGTCAAA 366
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 367 TTCACATCTACCTCTGTGATAACCTCAGTTATGGAACACAAACCTCTCTGTGCCAGTCA 426
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 427 CAGACCTCTGTATACGACAGTGTTCACCCAGCCCAACGTTTCAACTCCAGAGACA 486
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 487 ACCTTGAAGCCTAGCCTGTACCTCGAAATGTTTTCAGACCTTTCAACCACTAGCACTAGC 546
QY 141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys 160
Db 547 CTTGGCAACATCTCCCACTAAACCTTACATCATCTCTCTATCTATCTAAGTAGACATCAAG 606
QY 161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu 180
Db 607 GCAGAAATCAAAATGTTTCAGGCATCAGAGAAGTGAATTGACTCAGGGCATCTCCCTGGAG 666
QY 181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg 200
Db 667 CAAAATAAGACCTCCAGCTGTGGGAGTTTAAAGAAGACAGAGGGAGGGCCCTGGCCCGA 726
QY 201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu 220
Db 727 GTCTGTGTGGGGAGGAGGAGCTGATGCTGCTGCTGGGGCCAGGATGCTCCCTGCCTC 786
QY 221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu 240
Db 787 CTTGCCAGCTCTCAGGTGAGGCCCTCAGTGTCTACTGCTGTGCTGGCCACACAGAA 846
QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle 260
Db 847 ATTTCCAGCAAACTCCAACCTATGAAAGACACCAATCTGACCTGAAAGAGCTGGGGATC 906
QY 261 LeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle 280
Db 907 CTAGATTCTACTCAGCAAGATGTTGCAAGCCACAGACTATTCCCAAGAACCCCTGATT 966
QY 281 AlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet 300
Db 967 GCACCTGTACCTCGGAGCCCTGCTGGCTGTCTTGGGCATCATCTGCTATTTCCTCATG 1026
QY 301 AsnArgArgSerTrpSerProThrGlyGluArgLeu----- 312
Db 1027 AATCGCCGAGCTGGAGCCCCACAGAGAAAGGCT-GGAGCTGGAACCCCTGACCACCTCT 1085
QY 312 ----- 312
Db 1086 CAGGAAGAAAGGAGTCTGCACATGCGAGCTGCACCCCTCCCTCCGATCCTTCCCTCCACCTC 1145
QY 312 ----- 312
Db 1146 CCCCCTCCCTTCTCCACCCCTGCCCCACCTTCTCTGTTGGGCCCTCTCCCATCCAGTG 1205
QY 313 -----GlyGluAsp 315
Db 1206 TCTCAGAGCCCTGCTTACCAGATAAATGCTACTTTATTTATACATCTGCTAGGGCGAAGAC 1265
QY 316 ProTyrTyrThrGluAsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerPro 335
Db 1266 CCTATTACCGGAAACCGTGGAGCCAGGGCTATAGCTCAGGACCTGGGACCTCCCT 1325
QY 336 GluAlaGlnGlyLysAlaSerValAsnArgGlyAlaGlnGlnAsnGlyThrGlyGlnAla 355
Db 1326 GAGGCTCAGGAAAGCCAGCTGTGAACCGAGGGGCTCAGGAAACCGGAGCCGAGGCC 1385
QY 356 ThrSerArgAsnGlyHisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
Db 1386 ACCTCCAGAAACGGCCATTTCAGCAGACAAACACGCTGGTGGCTGATACCGAATTG 1439
```

```
RESULT 4
LOCUS S53911 2657 bp mRNA linear PRI 08-MAY-1993
DEFINITION CD34-glycoprotein expressed in lymphohematopoietic progenitor cells
(alternatively spliced, truncated form) [human, UT7, mRNA, 2657
nt].
ACCESSION S53911
VERSION S53911.1 GI:264768
KEYWORDS human UT7.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2657)
AUTHORS Nakamura,Y., Komano,H. and Nakauchi,H.
TITLE Two alternative forms of cDNA encoding CD34
JOURNAL Exp. Hematol. 21 (2), 236-242 (1993)
MEDLINE 93146100
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 124009] from the original journal article.
This sequence comes from Fig. 1AB.
FEATURES
source
1..2657
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene
1..2657
/gene="CD34"
CDS
91..1077
/gene="CD34"
/note="glycoprotein expressed in lymphohematopoietic
progenitor cells; This sequence comes from Fig. 1AB"
/codon_start=1
/product="CD34"
/protein_id="AAB25223.1"
/db_xref="GI:264769"
/translation="MLVRRGARAGPRMGWTAICLLSLPSGFMSLDNNGTATPELP
TQGTFSNVSTNVSYOETTPSTLGSTSLHPVSOHGNEATTNITETVKFTSVITSV
YGNNTSSVQSOTSVISTVETTPANVSTPETLKLPSLSPGNVSDLTSTSLATSPTKP
YTSSPILSDIAKEIKCGIRREVKLVTGICLQNKTSSCAEFKKRGRLARVLGEE
QADAGAGQVCSLLQAQSEVRPCILLVLANRTEISSKLQMKKHQSDLUKGLDFT
EDVASHQSYSQKTLIALVTGALLAVLIGITGYFLMNRSSWSPSTGERLELEP"
BASE COUNT 624 a 790 c 609 g 634 t
ORIGIN
Alignment Scores:
Pred. No.: 1.38e-110 Length: 2657
Score: 1827.00 Matches: 372
Percent Similarity: 84.93% Conservative: 0
Best Local Similarity: 84.93% Mismatches: 1
Query Match: 96.41% Indels: 66
DB: 9 Gaps: 1
US-09-836-602-2 (1-373) x S53911 (1-2657)
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 127 ATGCCCGGGCTGGACCGGCTTTCCTTCTGAGTTTGTGCTTCTGGTTTCATGAGT 186
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 187 CTTGACACACACGGTACTGCTACCCAGAGATTACCTACCCAGGAACATTTTCAATGTT 246
QY 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 247 TCTACAAATGATCTCTACCAAGAAACTACAAACACCTAGTACCTTGGAAAGTACCACTG 306
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrAsnIleThrGluThrValLys 80
Db 307 CACCCTGTCTCAACATGGCAATGAGGCCACAAACATCACAGAAACGACAGTCAAA 366
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
```


367	TTCACATCTACCTCTGTGTATTAACCTCAGTTTATGGAAACACAAACTCTCTGTGTCAGTCA	426
Qy	GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr	120
Db	427 CAGACCTCTGTATATCAGCACAGTCTTACCACCCAGCCAACTTCACTCCAGAGACA	486
Qy	121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer	140
Db	487 ACCTTGAAGCCCTAGCTCTCACTTGGAAATGTTTCAGACCTTTCAACCACTAGCACTAGC	546
Qy	141 LeuAlaThrSerProThrLysProTyrThrSerSerSerProIleLeuSerAspIleLys	160
Db	547 CTTCGAACATCTCCCACTAAACCTATACATCATCTTCTCTATTCCTAAGTGACATCAAG	606
Qy	161 AlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlu	180
Db	607 GCAGAAATCAATGTTCAGGCATCAGAGAGTGAATTGACTCAGGGCATCTGCCTGGAG	666
Qy	181 GlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArg	200
Db	667 CAAAATAAGACCTCCAGCTGTGCGAGTTTAAGAAGACAGGGGAGAGGGCTGGCCCGA	726
Qy	201 ValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeu	220
Db	727 GTGCTGTGTGGGAGGAGCAGGCTGATGCTGATCTGGGGCCCAAGGTATGCTCCTGCTC	786
Qy	221 LeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGlu	240
Db	787 CTTGGCCAGTCTGAGGTGAGGCCCTCAGTGTCTACTGTGCTTGGCCACAGAACAGAA	846
Qy	241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyIle	260
Db	847 ATTTCCAGCAAACTCCAACCTTATGAAAAGACCAACTCTGACCTGAAAAGCTGGGATC	906
Qy	261 LeuaspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIle	280
Db	907 CTAGATTTCACTGAGCAGAGATGTTGCAAGCCACAGAGCTATTTCCTCCAAAAGACCTGATT	966
Qy	281 AlaLeuValThrSerGlyAlaLeuAlaValLeuGlyIleThrGlyTyrPheLeuMet	300
Db	967 GCATGTGTCACTCGGAGGCCCTGCTGCTGCTCTTGGGCATCACTGGCTATTCTCTGATG	1026
Qy	301 AsnArgArgSerTyrSerProThrGlyGluArgLeu	312
Db	1027 AATGCCGCAGCTGGAGCCCCACAGGAGAAAGCT - GGAGCTGGGAACCTCACCACCTCTT	1085
Qy	312 -----	312
Db	1086 CAGGAAGAAAGAGTCTGCACATCAGCTGCACCTCCCTCCGATCCTTCTCCCACTC	1145
Qy	312 -----	312
Db	1146 CCCCTCCCTCTCCCAACCCCTGCCCCACTTCTCTGTTGGGCCCTCTCCATCCAGTG	1205
Qy	313 -----	315
Db	1206 TCTCACAGCCCTGCTTACCAGATAATGCTACTTTATTTATATACATGCTAGGGCGGAAGAC	1265
Qy	316 ProTyrTyrThrGluAsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerPro	335
Db	1266 CCTTATTACCGGAAACGGTGGAGCCAGGGCTATAGCTCAGGACCTGGGACCTCCCT	1325
Qy	336 GluAlaGlnGlyLysAlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAla	355
Db	1326 GAGGCTCAGGAAAGGCCAGTGTGAACGAGGGCTCAGGAAACAGGACCGCGCAGGCC	1385
Qy	356 ThrSerArgAsnGlyHisSerAlaArgGlnHisValValAlaAspThrGluLeu	373
Db	1386 ACCTCCAGAAAGGCCATTTCAGCAAGACACACCTGGTGGCTGATACCGAATTG	1439
RESULT 5		
CFU49457		
LOCUS	CFU49457	2956 bp
	mRNA	linear
		MAM 21-JAN-1997

```

DEFINITION  Canis familiaris hematopoietic progenitor cell marker CD34 mRNA,
complete cds.
ACCESSION  U49457
VERSION    U49457.1  GI:1224105
KEYWORDS   dog.
SOURCE     Canis familiaris
ORGANISM   Canis familiaris
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
AUTHORS    1 (bases 1 to 2956)
McSweeney,P.A., Rouleau,K.A., Storb,R., Bolles,L., Wallace,P.M.,
Beauchamp,M., Krizan-Benges,L., Moore,P., Sale,G., Sandmaier,B.,
de Revel,T., Appelbaum,F.R. and Nash,R.A.
TITLE      Canine CD34: cloning of the cDNA and evaluation of an antiserum to
recombinant protein
JOURNAL    Blood 88 (6), 1992-2003 (1996)
MEDLINE    96420219
REFERENCE  2 (bases 1 to 2956)
McSweeney,P.A.
AUTHORS    Direct Submission
TITLE      Submitted (16-FEB-1996) P.A. McSweeney, Transplant Biology, FHCR,
JOURNAL    1124 Columbia St, Seattle, WA 98104, USA
FEATURES   Location/Qualifiers
            source
            1..2956
               /organism="Canis familiaris"
               /db_xref="taxon:9615"
               /tissue_type="ML2 myelomonocytic leukemia"
            gene
            1..2956
               /gene="CD34"
            CDS
            260..1429
               /gene="CD34"
               /note="type 1 transmembrane protein"
               /codon_start=1
               /product="hematopoietic progenitor cell marker CD34"
               /protein_id="AAB41055.1"
               /db_xref="GI:1224106"
               /translation="MLAGRAGAGGLPRGWTALCLLSLPFGTINTETVITPTTPT
               STEMSAVSNTSKRAITLTPGTTTLYSVSDSGTTATISETTIVHVTSTSEITLT
               PCPNMSVQSOTSIATVTSPTNFSTSVTLPEPLLPGNGSDPYNSTSLVTSPTPEY
               YTSLSPTPSRNDPSTIKGIKSGYKVLNOGICLNETSCDEDFKDNKEKLTO
               VLCEPAEAGAVCSLLLAQSEVRPHCLLLVANKLETFSKLQLLRKHOSDLKKLGI
               RTEEDVGHQSHYSRIKTLAVTSGLLAVLTGTGYFLMNRSSWPTGERLGEDPY
               TENGGGGYSGSGPVSPEAQGKASVNRGPOENGCTGQATSRNGHSRQHVMVADTEL"
BASE COUNT 625 a 864 c 815 g 652 t
ORIGIN
Alignment Scores:
Pred. NO.: 3.28e-74 Length: 2956
Score: 1272.00 Matches: 265
Percent Similarity: 77.04% Conservative: 27
Best Local Similarity: 69.92% Mismatches: 79
Query Match: 67.12% Indels: 8
DB: 4 Gaps: 2

US-09-836-602-2 (1-373) x CFU49457 (1-2956)

Qy 1 MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 296 CTGCCGGGGCTGGACCGCGCTCTGCGTGTCTAGTCGTGCGCCCTTTGGGTTTCAAAAC 355

Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 356 ACAGAAACCGTGATTACTCTCTACCACAGTGCACACCTCCACAGAAATAATGTCAGCTGT 415

Qy 41 SerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSerLeu 60
Db 416 TCTGAGAATACATCCAAACGGGAAGCCATCACTAACTCTCTTGGAACTACCACTG 475

Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValValys 80
Db 476 TACTGTGCTCTCAAGACAGCAGGGGGACCACGACCACTCTCAGAGACTACAGCTCAT 535

Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100

```

Db 536 GTACATCTACCTCTGAGATCACCTTAACGCTGGGACCATGAACCTCTTCTTTCAGTCG 595
Qy 101 GlnThrSerValIleSerThrValPheThrProAlaAsnValSerThrProGluThr 120
Db 596 CAGACCTCTTACGCTATCAGGTATCTTTTACCCCAACCACTTTTCACTTCAAGTGTG 655
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 656 ACCTTGGAGCCAGCTGCTACCTCGAATGTTGGATCCCCCTACACAGCACCGAGC 715
Qy 141 LeuAlaThrSerProThrLysProThrThrSerSerPro----- 154
Db 716 CTTGTGACATCCCCACGGAATATTATACATCACTTCTCTACCCCAAGTAGAAATGAC 775
Qy 155 IleLeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThr 174
Db 776 ACCCAAGTACCATCAAGGGAGAAATCAATGTTCCGGAGTCAAGAAGTGAATGAAC 835
Qy 175 GlnGlyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArg 194
Db 836 CAAGGTATCTGCCTAGAGCTAAATGAGACCTCCACCTGTGAGGACTTTAAGAAAGATAAC 895
Qy 195 GlyGluGlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAla 214
Db 896 GAAGAAAAAAGTGAACCAAGTCTGTGT-----GAGAAAGAGCCACGCTGAGGCTGGGGCC 949
Qy 215 GlnValCysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuVal 234
Db 950 GGGGTGTCTCCCTGCTCTGCGCCAGTCTGAGGTGAGGCTTCACCTGCTGCTGTGTC 1009
Qy 235 LeuAlaAsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAsp 254
Db 1010 TTGGCCCAACAAACAGAACTTTTCAGTAACTCCAACTTCTCAGAAAGCACCACTCTGAC 1069
Qy 255 LeuLysLysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyr 274
Db 1070 CTGAAAAAGCTGGGGATCCGACATTCATCGAACAAAGATGTTGGAGCCACAGAGCTAT 1129
Qy 275 SerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIle 294
Db 1130 TCCGCCAAGACCTGATGTCATGTCACCTCAGGATGCTGCTGGCTGTCTTGGGCCACC 1189
Qy 295 ThrGlyTyrPheLeuMetAsnArgSerTrpSerProThrGlyGluArgLeuGlyGlu 314
Db 1190 ACTGTTACTTCTGATGACCGCGCAGTGGAGCCCTACAGGAGAAAGCTGGCGGAA 1249
Qy 315 AspProTyrTyrThrGluAsnGlyGlyGlyGlnGlyTyrSerSerGlyProGlyThrSer 334
Db 1250 GACCCTTATTACAGGAGAACGGTGGAGGCCAGGGCTATAGTCAAGGCCCTGGGGTCTCC 1309
Qy 335 ProGluAlaGlnGlyLysAlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGln 354
Db 1310 CTTGAGGCTCAGGGAAGCCAGTGTGAACCTGGGCTCAGGAGAACGGACGGCCGACG 1369
Qy 355 AlaThrSerArgAsnGlyHisSerAlaArgGlnHisValAlaAspThrGluLeu 373
Db 1370 GCCAGTCCAGAAAGGGCCATTACAGACAGACACACATGGTGGCTGTATACAGAAATTG 1426
RESULT 6
AF461503
LOCUS AF461503 2690 bp mRNA linear MAM 24-JAN-2002
DEFINITION Sus scrofa CD34 antigen mRNA, complete cds.
ACCESSION AF461503
VERSION AF461503.1 GI:18308137
KEYWORDS
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 2690)
Sun, J., and Butler, J.E.
Cloning, sequencing and expression of swine CD34 gene

Unpublished
2 (bases 1 to 2690)
Sun, J., Wang, J., and Butler, J.E.
Direct Submission
Submitted (20-DEC-2001) Microbiology, University of Iowa, 51 Newton
Road, Iowa City, IA 52242, USA
Location/Qualifiers
source
1. .2690
/organism="Sus scrofa"
/db_xref="taxon:9823"
/tissue_type="mesenteric lymph nodes"
298. .1467
/note="cluster of differentiation antigen 34"
/codon_start=1
/product="CD34 antigen"
/protein_id="AA167838.1"
/db_xref="GI:18308138"
/translation="MPRGWTTCLLSLLPSGFTAVNSTVASTLPTAAGSTPTGPGATA
TSAITLVPEVTWNSVQPTSLATVSSATINFTTSEVTLQPTSPGNTVAISGPTVNFMS
RSPSPVTSPPPTGSHKGEVKCAQIKVEKLTOGICLERNETSCERKFKKDNGBKLMO
ILCGEQEAERGPGVCSLLLAGSEVKPHCLLLVLANGETLSKFLLEKHQSELRMSI
QHFSKQDVRHQSYSRKTLTALVTSGILLAVLGTGTYLLMNRHSWSPTGRLGDPYI
TENGGGGQYSSGPGASPEAGKASVTRGAQENGTTGQATSRNGHSAROPMVADEL."
BASE COUNT 547 a 808 c 738 g 597 t
ORIGIN
Alignment Scores:
Pred. No.: 2.44e-70 Length: 2690
Score: 1212.00 Matches: 255
Percent Similarity: 73.60% Conservativeness: 35
Best Local Similarity: 64.72% Mismatches: 78
Query Match: 63.96% Indels: 26
DB: 4 Gaps: 6
US-09-836-602-2 (1-373) x AF461503 (1-2690)
Qy 1 MetProArgGlyTyrThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 298 ATGCCCGGGGCTGGACCCAGCTCTGCTTGTGCTGAGTTTGTCTGCTCTGCGCTTCACTACCT 357
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuPro----- 32
Db 358 GTG---AACAGCTCGACTGTTGCTTCCACCTTGCCAACTGGCGTGGCTCAACTCCCACC 414
Qy 33 -----ThrGlnGlyThr-----PheSerAsn 39
Db 415 GGGCCAGCTACTGCAGGGACAGCTATCACCGGTCAGTATCTCAGACATATCTTCACT 474
Qy 40 ValSerThrAsnValSerTyrGlnGluThrThrThrProSerThrLeuGlySerThrSer 59
Db 475 GTTCTCAATAATATATATCCACAGAGAA---ACCACACACAGATGCTTTCGAAGTCCAGC 531
Qy 60 LeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVal 79
Db 532 CTCCACACTGCTCTCTCAGGCGCAGTGGGACCCAGCCAGTACCTCTCAGGCCCTACACTT 591
Qy 80 LysPheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGln 99
Db 592 AATTTCATGTCTACCTCGCGGCTCACCTCTCCCGCAACCGTGAACCTCTCTGTCTCCAG 651
Qy 100 SerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGlu 119
Db 652 CCTCAGACTCT---CTAGCCACAGTGTCTCCGCCACCATCACTTTACAACTTCAGAG 708
Qy 120 ThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThr 139
Db 709 GTGACCTTCAGCCAGCAGCAGCTTCCAGAGAAATGTTTTCAGACCCCTCTACAAACAGTACC 768
Qy 140 SerLeuAlaThrSerProThrLysProThrThrSerSerSerProIleLeuSerAspIle 159
Db 769 AGCCCTGCGAGATCCCCACACGCCCCCTACACATCATCTCTCTCTACCCAGGTAGCCAC 828

```
Qy 160 LysAlaGluLeuLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeu 179
Db 829 AAGGGGAAGTCAATGTGCCAATCAAGAGGTGAATTCACCAAGGTATCGCTG 888
Qy 180 GluGlnAsnLysThrSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla 199
Db 889 GAGCGAAATGAGACCTCCGGCTGCGAAGAGTTTAAGAAGGACAATCGAGAGAAGTTGATG 948
Qy 200 ArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeu 219
Db 949 CAATTCCTGTGTGGCAGAGCAGCTGAGCGGGGCCAGG-----GTGTGCTCCCTTG 1002
Qy 220 LeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValValLeuAlaAsnArgThr 239
Db 1003 CTCCTTGCCCAATCTGAGTGAAACCTCAGTCGCTGCTGTGCTTGGCCAAACGGAACA 1062
Qy 240 GluLeuSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLysLysLeuGly 259
Db 1063 GAACCTAGCAGCAAGTTCTGCTGCTTCTGGAAGAACCAACAGCTCTGAAGTACGAGAGATGAGC 1122
Qy 260 IleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeu 279
Db 1123 ATCCACACTCTCGAACAAGATGTTAGGAGCCACAGAGACTCTCCGGAAGACCTTG 1182
Qy 280 IleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeu 299
Db 1183 ATTGCACCTGTCACCTCGGGATCCTGCTGCTGCTTGGGATCAGCTGCTACTGCTG 1242
Qy 300 MetAsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyrThr 319
Db 1243 ATGAACCGTCGAGTGGAGCCTACAGGAGAAGGCTGGGGAAGACCTTATTACAGC 1302
Qy 320 GluAsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGly 339
Db 1303 GAGACGGTGGAGCGAGGCTATTAGTCAGGCCCTGGGGCCCTCCCTGAGGCTCAGGGA 1362
Qy 340 LysAlaSerValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsn 359
Db 1363 AAAGCCAGTGTACTCAGAGGGCTCAGGAGAAGCGCACCGGCCACCTCCAGAAAC 1422
Qy 360 GlyHisSerAlaArgGlnHisValValAlaAspThrGluLeu 373
Db 1423 GGCCATTACGAGGCAACCCCATGTGTGGCAGATACTGAAGT 1464

RESULT 7
S69299
LOCUS S69299 1149 bp mRNA linear ROD 03-JUN-1994
DEFINITION CD34-cell surface antigen [alternatively spliced, clone 6-2-2]
ACCESSION S69299
VERSION S69299
KEYWORDS S69299.1 GI:495713
SOURCE Mus sp. stromal cell line PA-6.
ORGANISM Mus sp.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1149)
TITLE Suda, J., Sudo, T., Ito, M., Ohno, N., Yamaguchi, Y. and Suda, T.
JOURNAL Two types of murine CD34 mRNA generated by alternative splicing
MEDLINE Blood 79 (9), 2288-2295 (1992)
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 145456] from the original journal article.
This sequence comes from Fig. 3.
FEATURES
source
1..1149
/organism="Mus sp."
/db_xref="taxon:10095"
gene
1..1149
/partial
/gene="CD34"
CDS
1..1149
/gene="CD34"
```

```
/note="cell surface antigen; This sequence comes from Fig.
3"
/codon_start=1
/protein_id="AAB22108.1"
/product="CD34"
/db_xref="GI:495714"
/translation="MOVHRTFRAGLLLPWRWVALCLMSLLHNLNLTSTATTSTQGIS
PSVTNSESVENITSSIPGSTSHYLIYDSSKTTTPAISTVMVNTVTSIGPSGSGTTPH
TFSPQSTPTGILPTSDSISTSEMTWKSLLPSINVSDYSPNNSSFEMTSPTEPVAYT
SSASPAIKGEIKCSGIREVRLAQLCLELSEASCEFEKKEQSDLDLIQLCEKEARE
ADAGASCSLLUAQSEVRPECLMLVLANSTELPSKLQMKHKGDLRLGLQSFNKQD
IGSHQSYRSRKTALIALVTSGVLLATGLTGYELMNRNWSPTGERLGEDPYTENGQGG
GYSGPGASPETQKANVTGRGAQNGTGQATSRNGHSARQHVAVDEL"
BASE COUNT 307 a 293 c 285 g 264 t
ORIGIN
Alignment Scores:
Pred. No.: 2.66e-68 Length: 1149
Score: 1173.50 Matches: 245
Percent Similarity: 75.86% Conservative: 41
Best Local Similarity: 64.99% Mismatches: 80
Query Match: 61.93% Indels: 11
DB: 10 Gaps: 4
US-09-836-602-2 (1-373) x S69299 (1-1149)
Qy 1 MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer 20
Db 37 CTGCCATGGCGTGGGTAGCTCTCGCTGATGAGTCTGCTG-----CAT 81
Qy 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 82 CTAATAACTTGCATCTCTACACGAGACTTCTACACAGGAATATCCCATCAGTT 141
Qy 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 142 CCTACCAATGAGTCTGTTGAGAAATAATACATCATGATCCCTCGGAAGTACCAAC 201
Qy 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
Db 202 TACTTGATCTATCAGACAGCAGTAGACCAACACAGCCATCTCAGACATATGTTGTCAC 261
Qy 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 262 TTTACAGTTTACCTCTGGGATCCTTCAGGCTCTGGAACCTCCACACACTTTTTCACAACA 321
Qy 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 322 CAGACTTCCCAACTGGCATACTACTTCTCAGACAGATATTTCCTCAGAGATG 381
Qy 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSer 140
Db 382 ACTGGAAGTCCAGCTGCCATCTATAATGTTCTTGATATTTCGCTAATATAGCAGC 441
Qy 141 Leu---AlaThrSerProThrLysProTyr-----ThrSerSerProIleLeu 156
Db 442 TTTGAGATGACATCACCACCGAGCATATGCTTACACATCATCTTCTGCTCG----- 495
Qy 157 SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
Db 496 AGTGCCATTAAAGGGAGAAATCAATGCTCTGGAATCCGAGAAGTGAGGTGGCCCCAGGT 555
Qy 177 IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
Db 556 ATCTGCTGGAACCTAAGTGAAGCACTAGTTGTGAGGAGTTTAAGAAGAAAGGAGAA 615
Qy 197 GlyLeuAlaArgValLeuCysGlyGluGlnGlnAlaAspAlaAspAlaGlyAlaGlnVal 216
Db 616 GATCTAATTCAAATACTGTGTGAAGAGGAGGAGGCTGAGGCTGATGCTGTGCTAGTGTC 675
Qy 217 CysSerLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
Db 676 TGCTCCCTGCTTCTGACCCAGTCTGAGGTAGGCTGAGTGTGCTGATGCTGTGGCC 735
```



```

Db 1067 CCTGAGGCTCCCGGAGGCTCAGGAAAGGCCATTGTGAATCAGGGGCTCAGGAGAA 1126
Qy 351 GlyThrGluGlnAlaThrSerArgAsnGlyHisSerAlaArgGlnHisValValAlaAsp 370
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1127 GGCACCGCCAGGACCTCCAGAAACGCCATTACGCGAGAACACGTGTGGTGCCTGAT 1186
Qy 371 ThrGluLeu 373
||||| |||||
Db 1187 ACCGAAGT 1195

RESULT 11
AF259378
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Capra hircus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Caprinae; Capra.
1 (bases 1 to 2420)
Leipprandt, J.R., Anson, D.S. and Jones, M.Z.
Cloning and sequence analysis of caprine CD34 cDNA
Unpublished
2 (bases 1 to 2420)
Leipprandt, J.R., Anson, D.S. and Jones, M.Z.
Direct Submission
Submitted (20-APR-2000) Biochemistry, Michigan State University,
Room 519 Biochemistry, East Lansing, MI 48824, USA
FEATURES
source
1..2420
/organism="Capra hircus"
/db_xref="taxon:9925"
<1..1069
/feature="cell-surface antigen"
/codon_start=2
/product="CD34"
/protein_id="AAP76145.1"
/db_xref="GI:8515726"
/translation="PSMEFTNAVGTGNATVSSNPGISSPPTTTSYWKMSNSTFRST
SLYNSDSNGTAVTSVPASTVNFSTASSETLPASGVNSVOPQTSLATASSTLIS
FATSEATLOPSTPGNISDSLNTSPVTSINTSPSFTONILKTEIKCSRKEVK
LAHSICLKNETSNCEFEKKNNEKLTRVLCQEQAEVCSLLIAQSEVKPQCLLLVLT
NRTESKILRLKQSDLRMGIDISEEDVSHQYSRKRLIALTVLGLLAILIT
TCVFLNRRSWSFAGERLGEDPPYVENGSGGQYNSGSEASPEAQGKAIVNRGAQENGT
QOATSRNGHSARQVVDTEL"
BASE COUNT 565 a 700 c 590 g 565 t
ORIGIN

Alignment Scores:
Pred. No.: 8.38e-58 Length: 2420
Score: 1019.50 Matches: 216
Percent Similarity: 72.36% Conservative: 38
Best Local Similarity: 61.54% Mismatches: 88
Query Match: 53.80% Indels: 9
Db: 4 Gaps: 3

US-09-836-602-2 (1-373) x AF259378 (1-2420)

Qy 25 GlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnValSerThrAsnVal 44
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 35 GGTAAATGCCAGAGTGCA---TCGAACCCAGGAATATCTTCACCTGCCCTCAACACTACG 91
Qy 45 SerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeuHisProValSer 64
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 92 TCCCTACTGGAAGACATGAAGTCAAGTACTTTTGAAGACACACGAGCTCTACAATGTCTCT 151
Qy 65 Gln-----HisGlyAsnGluAlaThrThrAsnIleThrGluThrValLysPheThr 82
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152 CAGGACAGCAATGGGACACGACGATCACTCAGTTCCAGGCTCTACAGTCAATTTTCAG 211

```

```

Qy 83 SerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValcInSerGlnThr 102
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 212 TCTGCCCTGAGAGCCCTCCAGCCTCTGGAGCGGTGAAGTCTTCTGTCTCAACACGAGACC 271
Qy 103 SerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThrThrLeu 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 272 TCTTTAGCACACGAGGATCTTCTACCTCATCAGCTTTGGGACTTCAGAGCGGACTCTG 331
Qy 123 LysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSerThrSerLeuAla 142
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 332 CAGCCACGACGTTTCTGGAAATATTTCCGATCCCTATACATAGTACACGAGCCCTGTG 391
Qy 143 ThrSerProThrLysProTyrThrSerSerProIleLeuSerAspIleLysAlaGlu 162
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 392 ACTTCTTCCATTAATACCTCTCCATCATTTCTCTACCCAAAATATCTCAAGACTGAA 451
Qy 163 IleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeuGlnAsn 182
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 452 ATAAATGTTCCAGAGTCAAAGAGTGAATTTGAGGAGTGAAGTGAAGTGAAGTGAAGT 511
Qy 183 LysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArgValLeu 202
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 512 GAGACCTCCAACTGTGAGGAGTTTAAGAGAGAACAAATGAAGAGAGTGAAGTGAAGT 571
Qy 203 CysGlyGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeuAla 222
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 572 TGTCAAGAGGAGAACAG-----GCCAGGTGTGCTCTCTGCTCTGCTG 613
Qy 223 GlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThrGluLeuSer 242
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 614 CAGTCTGAGGTGAAGCCTCAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 673
Qy 243 SerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeuGlyLeuLeuAsp 262
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 674 AGCAAGATCAGACTTCTGAAAGGTACCAGTCTGACCTGAGAGAGATGGGATCCAGGAC 733
Qy 263 PheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeuIleAlaLeu 282
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 734 ATCTCTGAAGAAGATGTCCAGCACCAGAGCTATTCCCGGAAGACCTTGATTGCACTG 793
Qy 283 ValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeuMetAsnArg 302
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 794 GTCACCTCGGGGATCCTGCTGCGCATCTGATCACCAGCTGCTATTCTGTGATGAACGCG 853
Qy 303 ArgSerTrpSerProThrGlyGluArgLeuGlyGluAspProTyrTyrThrGluAsnGly 322
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 854 CGCAGCTGGAGCCCGACGAGAGAAAGGCTGGCGGAGACCCCTTATTACGTGGAGAACGGT 913
Qy 323 GlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLysAlaSer 342
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 914 GGAGCCAGGCGTATACTAGGCTCTGAGGCTCTCCCGAGGCTCAGGGAAGGCCATT 973
Qy 343 ValAsnArgGlyAlaGlnGlnAsnGlyThrGlyGlnAlaThrSerArgAsnGlyHisSer 362
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 974 GTGAATCGAGGGGCTCAGGAGAAATGGCAGCGCCAGGACACCTCCAGAAACGGCCATTCA 1033
Qy 363 AlaArgGlnHisValValAlaAspThrGluLeu 373
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1034 GCGAGACAAACGTGTGGTACGTATACCGAAGT 1066

RESULT 12
AF202879
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Homo sapiens hematopoietic progenitor cell antigen CD34 precursor
(CD34) mRNA, partial cds.
AF202879
AF202879.1 GI:6503195
human.
REFERENCE
1 (bases 1 to 581)

```


AUTHORS	Kutlar,F., Brisko,J., Leithner,C. and Kutlar,A.
TITLE	Direct isolation and cDNA sequencing of mRNA from human hematopoietic progenitor cell antigen CD34
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 581)
AUTHORS	Kutlar,F., Brisko,J., Leithner,C. and Kutlar,A.
TITLE	Direct Submission
JOURNAL	Submitted (08-NOV-1999) Medicine, Hematology and Oncology, Sickie Cell Center, Medical College of Georgia, 15th Street, AC-1000, Augusta, GA 30912, USA
FEATURES	Location/Qualifiers
source	1..581
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="1"
gene	/map="1q32"
	<1..>581
CDS	/gene="CD34"
	<1..>581
	/gene="CD34"
	/codon_start=2
	/product="hematopoietic progenitor cell antigen CD34 precursor"
	/protein_id="AAF14634.1"
	/db_xref="GI:65031136"
	/translation="ILSDIKAEIKCSGIREVKLTQIGICLEQNTSSCAFFKKDRGSLK ARVLCGEQADADAGQVCSLLIAQSEVRPQCLLVLANRTISSKLQIMKKHQSLK KIGILDFQDVASHQSYQSKTIALVTSGALLAVLGITGYFLMNRSSWSPGTERLGE D PYYTENGSGQGYSSGPGTSPRAQGRASVNRGA"
BASE COUNT	150 a 146 c 171 g 114 t
ORIGIN	
Alignment Scores:	
Pred. No.:	5,01e-56 Length: 581
Score:	980.00 Matches: 193
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	51.72% Indels: 0
DB:	9 Gaps: 0
US-09-836-602-2 (1-373) x AF202879 (1-581)	
QY 155	IleLeuSerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThr 174
Db 2	ATCTAAGTGACATCAAGGCAGAAATCAAATGTTTCAGGCATCAGAGAAGTGAAATTGACT 61
QY 175	GlnclyIleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArg 194
Db 62	CAGGGCATCTGCCTGGAGCAAAATAAGACCTCCAGCTGTGCGGAGTTTAAGAGGACAG 121
QY 195	GlyGluGlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAlaGlyAla 214
Db 122	GGAGGGCCCTGGCCCGAGTGCTGTGTGGGGAGGAGCAGCTCATGCTGATGCTGGGGCC 181
QY 215	GlnValCysSerLeuLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuVal 234
Db 182	CAGGTATGCTCCCTGCTCCCTTGCCAGTCTGAGTGAGGCGCTCAGTGTCTACTGCTGCTC 241
QY 235	LeuAlaAsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAsp 254
Db 242	TTGGCCCAACAGACAGAAATTTCCAGCAAACTCCAATTTATGAAAAGCACCAATCTGAC 301
QY 255	LeuLysLysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyr 274
Db 302	CTGAAAAGCTGGGGATCTCTAGATTTTCACGTGACACAGATGTTGCAAGCCACCAAGCTAT 361
QY 275	SerGlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIle 294
Db 362	TCCCAAAAGACCCCTGATTTGACCTGGTCACTCGGGAGCCCTGCTGGCTGCTTGGGCATC 421
QY 295	ThrGlyTyrPheLeuMetAsnArgSerTyrSerProThrGlyGluArgLeuGlyGlu 314
Db 422	ACTGGCTATTCTCTGATGAATCCCGCAGCTGGAGCCCCACAGAGAAAGGCTTGGCGAA 481

Qy	315	AspProTyrThrGluAsnGlyGlyGlnGlyTyrSerSerGlyProGlyThrSer	334
Db	482	GACCCTTATTACAGGAAACGGTGGAGCGCAGGCTATAGCTCAGGACCTGGGACCTCC	541
Qy	335	ProGluAlaGlnGlyLysAlaSerValAsnArgGlyAla	347
Db	542	CCTGAGGCTCAGGAAAGCCAGTGTGAACCGAGGGGCT	580
RESULT 13			
LOCUS	AR078454	1065 bp	linear
DEFINITION	Sequence 18 from patent US 5962644.	DNA	PAT 31-AUG-2000
ACCESSION	AR078454		
VERSION	AR078454.1	GI:10005200	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 1065)		
AUTHORS	Hawley,R.J. and Monroy,R.L.		
TITLE	Porcine CD34		
JOURNAL	Patent: US 5962644-A 18 OCT-1999;		
FEATURES	Location/Qualifiers		
source	1..1065		
BASE COUNT	248 a 326 c 273 g	218 t	
ORIGIN	/organism="unknown"		
Alignment Scores:			
Pred. No.:	1.04e-50	Length:	1065
Score:	904.00	Matches:	196
Percent Similarity:	69.44%	Conservative:	38
Best Local Similarity:	58.16%	Mismatches:	77
Query Match:	47.70%	Indels:	26
DB:	6	Gaps:	6
US-09-836-602-2 (1-373) x AR078454 (1-1065)			
Qy	1	MetProArgGlyTrpThrAlaLeuCysLeuLeuSerLeuLeuProSerGlyPheMetSer	20
Db	67	ATGCGCGGGCTGGACGACGCTCGTGTGCTGAGTTGCTGCCCTCTGGGTTTCACAGCT	126
Qy	21	LeuAspAsnAsnGlyThrAlaThrProGluLeuPro	32
Db	127	GTG---AACAGCTCAACTATTGCTTCGACCTTGCACGCTGCCGCTGGGTCAACTCCGACC	183
Qy	33	-----ThrGlnGlyThrPheSerAsn-----	39
Db	184	GGCGCGGTACCGCAGGGCAGCTATCACCGGGTCACTATCTCAGACATATCTTCACCT	243
Qy	40	ValSerThrAsnValSerTyrglnGluThrThrThrProSerThrLeuGlySerThrSer	59
Db	244	GTTTCTCAATATATCCACGAGGAACACATCA---GATGCTTTCGAAAGTGCACG	300
Qy	60	LeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrThrVal	79
Db	301	CTCCACACTGTCCTCAGGCGCAGCGTGGGACCACTAGCCATCTCAGGCCCTACAGTT	360
Qy	80	LysPheThrSerThrSerValIleThrSerValTyrglyAsnThrAsnSerSerValcIn	99
Db	361	AATTTCATGCTACCTCGCGGGTCACCTCGTCCCGCAACCGTTAACTCTCTGTCTCAG	420
Qy	100	SerGlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGlu	119
Db	421	CCTCAGACCTCT---CTAGCCACACGCGTCCTCCGCCCACTCAACTTTTCAACTTCAG	477
Qy	120	ThrThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrSerThr	139
Db	478	GTGACCTTGCAGCCGACGAGTTCCTCCAGGAATGTTTCAGACCCCTCTACACAGTACC	537
Qy	140	SerLeuAlaThrSerProThrLysProTyThrSerSerProIleLeuSerAspIle	159


```
Db 538 AGCCTCGAGATCCCCACAGCCCTACACATCATCTCTCTACCCAGGTAGCCAC 597
QY LysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGlyIleCysLeu 179
Db 598 AAGGGGAAGTCAATGTGCCCCAAATCAAGAGGTGAATTTGACCCAGGATCTCGCTG 657
QY GluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAla 199
Db 658 GAGCGAATCAGACCTCCGCTCGGAGAGCTTTAAGAAGGACAAATGAGGAGAAGTTGATG 717
QY 200 ArgValLeuCysGlyGluGlnAlaAspAlaAspAlaGlnValCysSerLeu 219
Db 718 CAAATCTGTGTGGCAGGAGAGGCTGAGGCGCGCCAGGG-----GTGTGCTCCTTG 771
QY 220 LeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAlaAsnArgThr 239
Db 772 CTCCTGCCCAATCTGAGGTGAACCTCACTGCTGCTGTGCTGTGCTGTGGCCACGGAACA 831
QY 240 GluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLysLysLeuGly 259
Db 832 GAATCTAGCAGCAAGTTCCTGCTCTGGAAGAGCACCAGTCTGAACCTGAGAGATGAGC 891
QY 260 IleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGlnLysThrLeu 279
Db 892 ATCCAAACTTCTCGAAACAAAGATGTTAGGAGCCACCAGAGCTACTCCCGAAAGACCTTG 951
QY 280 IleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGlyTyrPheLeu 299
Db 952 ATTGCACTGTGCTACCTCGGGATCCTGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 1011
QY 300 MetAsnArgArgSerTrpSerProThrGlyGluArgLeuGlyGluAspPro 316
Db 1012 ATGAACCGTCGCAGTTGAGCCCTACAGAGAAAGGCTGGAGCTGGAAACC 1062

RESULT 14
S69301
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REMARK
FEATURES
source
gene
CDS
```

597 bp mRNA linear ROD 03-JUN-1994
CD34-cell surface antigen (alternatively spliced, clone 2) [mice,
stromal cell line PA-6, mRNA Partial, 978 nt].

S69301.1 GI:495715
Mus sp. stromal cell line PA-6.
Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 978)
Suda,J., Sudo,T., Ito,M., Ohno,N., Yamaguchi,Y. and Suda,T.
Two types of murine CD34 mRNA generated by alternative splicing
Blood 79 (9), 2288-2295 (1992)
92239883
GenBank staff at the National Library of Medicine created this
entry [NCBI gi15459] from the original journal article.
This sequence comes from Fig. 3.
Location/Qualifiers
1..978
/organism="Mus sp."
/db_xref="taxon:10095"
1..978
/partial
/gene="CD34"
1..978
/gene="CD34"
/note="cell surface antigen; This sequence comes from Fig.
4"
/codon_start=1
/product="CD34"
/protein_id="AAB22109.1"
/db_xref="GI:495716"
/translation="MOVHRTDTRAGLLPWRVALCLMSLLHLNLTSTATERSTQGIS
PSVPTNESVEENITSSPTSHYLYQSSKTPAISETMVNFTVTSIPSGSGTPH
TFSPQTSPTGLPTTSDISTSEMTWKSLSIPNSDYSPNNSSSEFTSPTPEFYAT
SSSAPSAIKGEIKCSGIREVRLAQICLSELSASSCEEFKKEKEDLIQICKEAE

```
ADAGASVCSLLLAQSEVRPECLLMVLANSTELPSKLQLMKHQSDLRKLGIOFNKQD
IGSHOSYSRKTLIALVTSGVLLAILCTGTYFLMNRSSWPTGERLEP"
BASE COUNT 261 a 249 c 228 g 240 t
ORIGIN
Alignment Scores:
Pred. No.: 1.44e-48 Length: 978
Score: 870.50 Matches: 189
Percent Similarity: 71.88% Conservative: 41
Best local Similarity: 59.06% Mismatches: 79
Query Match: 45.94% Indels: 11
DB: 10 Gaps: 4
US-09-836-602-2 (1-373) x S69301 (1-978)
QY 1 MetProArgGlyTrpThrAlaLeuCysLeuSerLeuLeuProSerGlyPheMetSer 20
Db 37 CTGCCATGGCGCTGGGTAGTCTCTGCTGATGATGCTGCTG-----CAT 81
QY 21 LeuAspAsnAsnGlyThrAlaThrProGluLeuProThrGlnGlyThrPheSerAsnVal 40
Db 82 CTAATTAACCTTGACTCTGCTACACGAGACTTCTACAAAGGAATATCCCAATCAGTT 141
QY 41 SerThrAsnValSerTyrGlnGluThrThrProSerThrLeuGlySerThrSerLeu 60
Db 142 CCTACCAATGAGTCTGTTGAGGAAATATCACATCTAGCATCTCCCTGGAAGTACCAGCAC 201
QY 61 HisProValSerGlnHisGlyAsnGluAlaThrThrAsnIleThrGluThrValLys 80
Db 202 TACTTGTATCTATCAGGACAGCAGTAAGACACACACAGCCATCTCAGAGACTATGGTCAAC 261
QY 81 PheThrSerThrSerValIleThrSerValTyrGlyAsnThrAsnSerSerValGlnSer 100
Db 262 TTTACATTTACCTCTGGGATCCCTTCAGGCTCTGGAACTCCACACACTTTTTCACACCA 321
QY 101 GlnThrSerValIleSerThrValPheThrThrProAlaAsnValSerThrProGluThr 120
Db 322 CAGACTTCCCACTGGCATACTGCCTACTTTCACACAGATATTCCACTTTCAGAGATG 381
QY 121 ThrLeuLysProSerLeuSerProGlyAsnValSerAspLeuSerThrThrSer 140
Db 382 ACCTGGAAATCCAGCCTGCCATCTATAAATGTTTCTGATTATTCGCCCTAATAATAGCAGC 441
QY 141 Leu---AlaThrSerProThrLysProTyr-----ThrSerSerSerProIleLeu 156
Db 442 TTTGAGATGACATCACCACCGAGCCATATGCTTACACATCATCTTCTGCTCCG----- 495
QY 157 SerAspIleLysAlaGluIleLysCysSerGlyIleArgGluValLysLeuThrGlnGly 176
Db 496 AGTGCCATTAAAGGAGAGAAATCAATGCTCTGGAATCCGAGAAGTGAGGTGGCCAGGGT 555
QY 177 IleCysLeuGluGlnAsnLysThrSerSerCysAlaGluPheLysLysAspArgGlyGlu 196
Db 556 ATCTGCCCTGAAACTTAAGTGAAGCATCTAGTTGTGAGGAGTTTAAGAGGAAAGGAGAA 615
QY 197 GlyLeuAlaArgValLeuCysGlyGluGluGlnAlaAspAlaAspAlaGlyAlaGlnVal 216
Db 616 GATCTAATTCATATCTGTGTGAAGAGGAGGAGGCTGAGGCTGATGCTGTGCTGATGTC 675
QY 217 CysSerLeuLeuLeuAlaGlnSerGluValArgProGlnCysLeuLeuValLeuAla 236
Db 676 TGCTCCCTGCTTCTAGCCAGTCTGAGGTAGGCTGAGGCTGAGTGTGCTGATGCTTGGCC 735
QY 237 AsnArgThrGluIleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLys 256
Db 736 AATAGCACAGAACTTCCCAAACTCCAGCTTATGGAAGAGCACCACCAATCTGACTTGAGA 795
QY 257 LysLeuGlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSerGln 276
Db 796 AAGCTGGGATCCAAAGCTTCAATAAACAAGATATCGGGAGCCACAGAGCTATTCCCGA 855
QY 277 LysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThrGly 296
```



```

/notes="AluY repeat: matches 1. .311 of consensus"
17989..18227
/notes="L2 repeat: matches 1995. .257 of consensus"
18250..18508
/notes="Charlie5 repeat: matches 2315. .2577 of consensus"
18510..18698
/notes="Charlie5 repeat: matches 16. .196 of consensus"
18721..18935
/notes="Charlie1b repeat: matches 3. .219 of consensus"
18936..19243
/notes="AluY repeat: matches 1. .307 of consensus"
19244..19531
/notes="Charlie1b repeat: matches 219. .501 of consensus"
19610..19779
/notes="L1MD repeat: matches 1. .169 of consensus"
19780..20228
/notes="L1MD repeat: matches 787. .1293 of consensus"
20299..20606
/notes="L2 repeat: matches 2382. .2710 of consensus"
20625..20753
/notes="L2 repeat: matches 2179. .2307 of consensus"
20813..20960
/notes="L2 repeat: matches 2374. .2523 of consensus"
21149..21292
/notes="MIR repeat: matches 110. .261 of consensus"
21334..22002
/notes="L2 repeat: matches 1947. .2724 of consensus"
<21938..22360
/notes="match: GSS AQ193501 clone 2383C13"
<22003..>22108
/notes="match: multiple ESTs; match: AA909597 AA906144
AA907722"
22573..22657
/notes="MIR repeat: matches 81. .160 of consensus"
complement(24084..25666)
/notes="match: STS G06650"
complement(join(24102..25486,26245..26409,26724..26776,
27028..27184,35057..35137,36529..36782,37377..37559,
48563..48899))
/genes="CD34"
/notes="match: M01104; X60172; match: multiple ESTs; match:
T80274 C18678 AA553071 AA311526 AA311473 AA907722 AA906144
AA064307 AI173145 AA369014 R39493 AA148360 T81557 R72478
R69703 R69740 AA039660 AA639555 R73090 R73955 W92532
R66845 AA054965 AA434483 R69214 R82715 AA188435 AA524276
W74409 H01096 T12007 AA216274 AA022916 AA528783 AA483672
AA640408 W58493 T28504 AI128488 AI149563 AI144193 AI252665
AI017002 AI160509 AI017793 AA906021 AI128776 AI150834
AA434387 AA664247 AA039661 W72884 AI262206"
evidence=not_experimental
complement(24102..48899)
/genes="CD34"
complement(join(25301..25486,26245..26409,26724..26776,
27028..27184,35057..35137,36529..36782,37377..37559,
48563..48605))
/genes="CD34"
/notes="match: SWISS-PROT:P28906"
/codon_start=1
/evidence=not_experimental
/product="CD34 (hematopoietic stem cell surface antigen)"
/protein_id="CAB36556.1"
/db_xref="GI:4455631"

```

Alignment Scores:

```

Pred. No.: 5.64e-30 Length: 160771
Score: 630.50 Matches: 185
Percent Similarity: 29.32% Conservative: 0
Best Local Similarity: 29.32% Mismatches: 4
Query Match: 33.27% Indels: 445
DB: 9 Gaps: 3

```

US-09-836-602-2 (1-373) x HS88L2 (1-160771)

```

QY 185 SerSerCysAlaGluPheLysLysAspArgGlyGluGlyLeuAlaArgValLeuCysGly 204
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 27193 TCTTCCAGCGGAGTTTAAAGAGGACAGGAGGCGCTGGCCCGAGTCTGTGGG 27134
QY 205 GluGluGlnAlaAspAlaAspAlaGlyAlaGlnValCysSerLeuLeuLeuAlaGlnSer 224
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 27133 GAGGAGCAGCTGATGCTGATGCTGGGCGCCAGGTATGCTCCCTGCTCTGCCAGTCT 27074
QY 225 GluValArgProGlnCysLeuLeuValLeuAlaAsnArgThr----- 239
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 27073 GAGGTGAGGCTCTAGTGTCTACTGCTGTGCTTGGCCCAACAAGAAC-AGGTAAGGTGCACCT 27015
QY 239 ----- 239
Db 27014 CTGCCTGGGGACAGGGAAGGAGTAGGGCTGAGGTTAGAGATCTCTGGGTGGAGATGGG 26955
QY 239 ----- 239
Db 26954 GCATCTTAGATCCAGAGAGACCACAGGTGCTGGGGAGAAGAGCTTGGCTGGCTTTGGGAG 26895
QY 239 ----- 239
Db 26894 CGTCCCGCCGAGATGGACCAACCCTGCCCATCAGAGCATCTTTAGAACACAGCATGGGAGG 26835
QY 240 -----Glu 240
|||||
Db 26834 GTGGAGCAGGATAAAGCTGGTTCTCTTTAGATTAGCAACCCCTGTTCTTAATTCGTAGAA 26775
QY 241 IleSerSerLysLeuGlnLeuMetLysLysHisGlnSerAspLeuLysLysLeu----- 258
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 26774 ATTTCCAGCAAACTCCAACCTATGAAAAAGCACCACCAATCTGACCTGAAAAAGGT-AAGTCC 26716
QY 258 ----- 258
Db 26715 TCTGAGATGATGGGCCTCGAGAGGGAATCCCGGGCAGGGGGCTCCCTGGAGTAAGGGGG 26656
QY 258 ----- 258
Db 26655 GAAGTCATTATCTTCCCCAGTTCTACCTACCATCTTCTCTGAGTCTTCTCCTTTCTTAGT 26596
QY 258 ----- 258
Db 26595 CTTGAACCAAAAATGGTGGGGTATGACTGAATTAAGAGAGCTGAGTCTCAAAATCCCATTT 26536
QY 258 ----- 258
Db 26535 CAAAAATTCCTGAGGTCAATGCCCTCAAGCATCTGTTTCACTGTTGAGTCACACACAGGGG 26476
QY 258 ----- 258
Db 26475 GCTGGAGCGGGCTGGGCTCCACCTTTTGGCAAGTCACAGGCTGTATCTTCTTGTCTTCTT 26416
QY 259 -----GlyIleLeuAspPheThrGluGlnAspValAlaSerHisGlnSerTyrSer 275
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 26415 TTCAAGCTGGGATCTCTAGATTTTCACTGAGCAAGATGTTGCAAGCCACCAGAGCTATTCC 26356
QY 276 GlnLysThrLeuIleAlaLeuValThrSerGlyAlaLeuLeuAlaValLeuGlyIleThr 295
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 26355 CAAAGACCTGATTTGACCTGGTCCACCTCGGAGCCCTGCTGGCTGTCTTGGGCATCACT 26296
QY 296 GlyTyrPheLeuMetAsnArgArgSerTrpSerProThrGlyGluArgLeu----- 312
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 26295 GGCTATTTCTGATGAATCGCCGAGCTGGAGCCCAACAGGAGGCT-GGTCAGTTC 26237
QY 312 ----- 312
Db 26236 TGGGGCCAGGGTAAAGAAATAGGAAGATAGTGGGTTCTCTGGGAGTTCAGTGGATGT 26177
QY 312 ----- 312
Db 26176 CATGGAGCAGGAGGAGAAATAGAAAAAGCCCTTCTGTGAGCTTACATAAAGATATGC 26117
QY 312 ----- 312

```

```
Db 26116 ATGTGTGCACACACAGTAATCGGTGAAGATTCAAAATATATTATGTAGCAAACTGGGGAGG 26057
QY 312 ----- 312
Db 26056 GGACAGTAATGGTGGCCAGCCACCTACTCGGTGTGTAGCAGATGATTCGTGTCACTCTAT 25997
QY 312 ----- 312
Db 25996 TCATTATGTTTTCTGGATAGTTCAGCTTGGGGCCACCAGAACAGGCTGTTTAATCAGCC 25937
QY 312 ----- 312
Db 25936 ACACCTGATTTGCAACCGTGTAAATGCCAGTGGGTGCCCTTGCTCAAAAGGAGGTA 25877
QY 312 ----- 312
Db 25876 TACGGAGGAGAAATCCCATTTGCCATTTCTGGATGAGGAGGACAAAGGCTGAGGTCTGAA 25817
QY 312 ----- 312
Db 25816 TCTTGGCCTCTGGCCTGTCCCTTACCCTGGGAGGTCTATCCACCCCTTCTTGGAACTGCC 25757
QY 312 ----- 312
Db 25756 CGTTTTCCCGGGAGCTGGCTGCCGAGCTGCGGTGGCATGGTGTGCCCTCTCCTATG 25697
QY 312 ----- 312
Db 25696 TCCTTTCTCCTTAGGAGCTGGAACCTGACCACCTTTCAGGAAGAAAGAGTCTGCACA 25637
QY 312 ----- 312
Db 25636 TGCAGCTGCACCCCTCCCTCCGATCCTTCTCCACCTCCCTCCCTTCTCCACCCC 25577
QY 312 ----- 312
Db 25576 TGCCCCCACTTCTGTGTTTGGGCCCTCTCCCATCCAGTGTCTCACAGCCCTGCTTACCAG 25517
QY 313 -----GlyGluAspProTyrThrGluAsnGly 322
Db 25516 ATAACTCTACTTTATTTATACACTGTCTAGGGCGAAGACCCCTATTATACACGGAAACGGT 25457
QY 323 GlyGlyGlnGlyTyrSerSerGlyProGlyThrSerProGluAlaGlnGlyLysAlaSer 342
Db 25456 GGAGGCCAGGGCTATAGCTCAGGACCTGGGACCTCCCTGAGGCTCAGGGAAGGCCAGT 25397
QY 343 ValAsnArgGlyAlaGlnGluAsnGlyThrGlyGlnAlaThrSerArgAsnGlyHisSer 362
Db 25396 GTCAACCGAGGGGCTCAGGAAACGGGACCGGCCAGCCCTCCAGAAACGGCCATTCA 25337
QY 363 AlaArgGlnHisValValAlaAspThrGluLeu 373
Db 25336 GCAAGACAACACGTGGTGGCTGATACCGGAATTG 25304
```

Search completed: October 30, 2002, 09:51:15
Job time : 1971.1 secs